KU Leuven
Group Biomedical Sciences
Faculty of Pharmaceutical Sciences
Rega Institute for Medical Research
Medicinal Chemistry

Vectors for the active uptake of nucleotides in bacterial cells - A model study for the delivery of XNA building blocks

Swarup DE

Doctoral thesis in Pharmaceutical Sciences

Leuven, 2015

KU Leuven
Group Biomedical Sciences
Faculty of Pharmaceutical Sciences
Rega Institute for Medical Research
Medicinal Chemistry

Vectors for the active uptake of nucleotides in bacterial cells - A model study for the delivery of XNA building blocks

Swarup DE

Jury:

Promotor: Prof. Piet Herdewijn
Co-promotor: Dr. Elisabetta Groaz
Chair: Prof. Arthur Van Aerschot
Secretary: Prof. Jef Rozenski
Jury members: Prof. Steven Ballet
Prof. Wim Dehaen
Prof. Abram Aertsen

Leuven, 02.07.2015

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Pharmaceutical Sciences

Dedicated To My Loving Family

Acknowledgements

This 4 years journey towards my dissertation would not have been possible without many wonderful people who have consciously or unconsciously supported me in many ways.

First and foremost, I would like to express my deep and sincere gratitude to my promoter ***Prof. Piet Herdewijn*** *for providing me the opportunity to carry out this research work in his group, for his scientific ideas, advice, fruitful discussions and constructive criticism, and for introducing me to the multidimensional field of synthetic biology. I am also indebted to my co-promotor* ***Dr.Elisabetta Groaz*** *for her constant help, support and guidance, especially for carefully reading and correcting all the manuscript in this journey.*

I thank our scientific and technical collaborators for their supports in our study: ***Prof. Abram Aertsen****,* ***Dr. Anirnan Ghosh, Dr. Valerie Pezo, Dr. Philippe Marliere*** *for providing the auxotrophic bacterial strains and* ***Lia Margamuljana*** *and* ***Dr. Elisabetta Groaz*** *for their support in carrying out biological experiments of my compounds.*
I would like to give my special thanks to all the jury members ***Prof. Arthur Van Aerschot, Prof. Jef Rozenski, Prof. Abram Aertsen, Prof. Steven Ballet, Prof. Wim Dehaen*** *for the time they have provided in carefully scrutinising my thesis and coming up with suggestions to improve the quality of the thesis. A special thanks to* ***Prof. Arthur Van Aerschot*** *for accepting to be the chair of my doctoral evaluation committee and for his scientific and technical advises.*

I thank ***Prof. Jef Rozenski*** *for mass spectrometry and LC/HR-MS measurements,* ***Prof. Mathy Froeyen*** *for molecular modelling studies. I would also like to thank* ***Prof. Eveline Lescrinier*** *for scientific help and discussions related to the 2D NMR assignments.*

I take the opportunity to thank all our secretaries ***Chantal Biernaux****,* ***Christiane Callebaut****,* ***Dominique Brabants****,* ***Inge Aerts,*** *and* ***Cathy De Meyer*** *for their excellent secretarial help, a special thanks to* ***Chantal*** *for her excellent editorial help for publishing journal articles.*

I am extremely thankful to our lab technicians ***Luc Baudemprez*** *for recording 500 and 600 MHz NMR spectra and for his technical assistance for arranging lab equipments,* ***Guy Schepers*** *for overall lab management.*

I also thank my colleague and only Belgian friend ***Bart Vandenbrande*** *for carefully translating my research summary into Dutch.*

This thesis will be incomplete without mentioning my teachers and professors who encouraged me and taught me chemistry and biology in college and institute. Especially I am indebted to my master

*degree project supervisor **Prof. Jayanta K. Ray** and **Dr. Sunil K. Ghosh** who helped and motivated me to grow as a synthetic organic chemist. I am also very grateful to my previous boss **Mrs. Ramya Rajan** from Syngenta who always encouraged me to take up any challenge in life.*

*I sincerely thank my friend and colleague **Dr. Mohitosh Maiti** for his initial guidance to 'Living in Leuven', helpful scientific discussions, encouragements, all-round knowledge and helping in learning the 2D NMR assignments. I also thank **Dr. Mikhail Abramov**, **Dr. Puneet Srivastava, Dr. Om Prakash Bande, Dr. Shrinivas Dumbre** for their scientific advice, help and support throughout this journey. Many thanks to **Puneet** for sharing his immense knowledge in various topics mostly during silent hours and **Om** for various constructive discussions in my leisure time during HPLC purification in the 6th floor.*

*I would like to thank all my medicinal chemistry (ex)lab mates **Steven, Natalia, Munmun**, **Amit, Nastia, Bart**, **Bharat**, **Chao**, **Elena**, **Fabio**, **Hassen**, **Li**, **Lia**, **Marleen**, **Milind**, **Miyeon**, **Min**, **Nastia**, **Qiuya**, **Rania**, **Marta**, **Song**, **Stefano, Venkatesham, Mengmeng, Baole, Piotr, Rahul, Elzbieta, Randy, Henri, Manesh, Marta** for being such a nice and supportive co-worker all the time.*

I thank Erasmus Mundus-Experts grants and KU Leuven research grants for providing financial support in this long journey.

*I do not have words for expressing my feelings for my parents and other family members for giving me so much of love, support and affection always. Without your support, I would not have been here today, in particular, my uncle cum teacher **Mr. Laxman Chandra De** who motivated me always from my childhood to take up Science for higher studies and brother cum tutor **Mr. Abhoy Patra** by profession now a Chemistry teacher, inspired me the most to study Chemistry as a major.*

Apart from professional life, my time at Leuven was made enjoyable in large due to many friends and groups that became a part of my social life, many thanks to you all for the wonderful homely environment which helped me keeping a balanced life and staying focused on the track.

*Last, but not least, many thanks to my beloved wife **Sonai** for her unconditional love, encouragement and moral support in every circumstance. I thank for her patience and understanding for not being with her in most of the weekends. My sweetest thanks to our seven-month-old daughter **Sohalia** for being a lovely girl and letting me submit this thesis on time.*

Swarup De

Leuven, 2015

Summary

The central dogma of molecular biology relies on the storage, transfer and propagation of genetic information from DNA to protein via RNA. Synthetic genes have been widely used to create genetically modified organisms, for a number of applications in therapeutics, agriculture, energy and environment related fields, material science and biotechnology. However, due to accidental or deliberate damage triggering from the exchange of genetic information with natural species via gene transfer, such artificial genetic constructs are generally considered as a potential threat to wild ecosystems.

The development of orthogonal biological systems, through the chemical assembly of nucleic acids which do not occur in nature (Xeno Nucleic Acids, XNAs), has been conceived as a powerful genetic firewall to avoid genetic cross talk, thus addressing crucially related biosafety issues. Several XNA based-systems have been synthesized with this purpose, following modifications at the base, sugar and leaving group moieties along with the engineering of enzymes showing XNA associated catalytic activities. In order to reach their true evolutionary potential, however, reprogrammed organisms require the presence of XNA nucleotidic precursors of XNA polymers in sufficient amount inside the cell. The *in vivo* invasion of XNA replicons was envisioned to firstly materialize in bacteria, as anticipated by their relatively simple genetic architecture. Although most modified nucleosides could easily enter bacterial cells by passive diffusion, they are known to be poorly phosphorylated to their mono-phosphate forms by wild-type kinases. On the other hand, the delivery of nucleotides is restricted by hydrophobic interactions with the lipid bi-layer of the bacterial cell wall, as a result of their low pKa and high polarity. Bacteria nonetheless possess several transmembrane transport proteins that facilitate uptake, against a concentration gradient, of nutrients for their survival, such as peptides, sulfonates, sugars and vitamins etc., which otherwise would experience limited entry.

On this basis, this doctoral thesis aims at exploring the viability of natural systems devoted to the transport of common metabolites as potential delivery tools for nucleotides in bacterial cells, by means of the systematic design and synthesis of different nucleotide-conjugates and their evaluation in a nutritional selection assay. Thymidine mono-phosphate (TMP) was chosen as exemplary nucleotide in a model study and the intracellular delivery of the corresponding TMP-conjugates could be established based on the growth of a suitable *E. coli* mutant auxotrophic for thymidine (thyA-, ΔphoA).

In the first project (**Chapter 2**), a variety of di- to penta-peptides were chosen as delivery vehicles targeting different peptide permeases, thus a series of peptide-TMP conjugates (PNCs) were synthesized to be evaluated for their functional activity in *E. coli*. A number of linear and convergent synthetic strategies along with orthogonal protection-deprotection approaches are discussed for the synthesis of such conjugates, obtained by connecting the side chain functionalities of oligopeptides to

the 3'-hydroxyl group of TMP through bio-cleavable linkers such as carbamate (OCONH), ester (OCO), oxyamide (ONHCO), oxymethyleneoxyamide (OCH_2ONHCO) and oxymethylene-oxyester (OCH_2OCO).

In **chapter 3**, design and synthesis of various PNCs of TMP are discussed, connected through various phosphoester linkers i.e. phosphate, phosphoramidate, phosphorothioate and thioethyl-phosphate via the 5'-hydroxyl of thymidine and the α-carbon of the C-terminal glycine residue of L-Ala-Gly. To gain a preliminary insight of potential *in vivo* metabolism, these conjugates were also evaluated as substrates in the template directed DNA-polymerization with different thermophilic and mesophylic DNA polymerases e.g. Therminator, Vent (exo-) and Klenow (exo-). A free carboxylic functionality is necessary for binding in the active site of the enzymes. The incorporation kinetic data (V_{max}/K_M) of the conjugate comprising a phosphoramidate linker and a free carboxylic group was found to be 994 fold lower than the natural substrate TTP, within the range of leaving groups mimics previously explored in our lab.

In **Chapter 4**, novel sulfono-phosphoramidate conjugates exhibiting at 5′-phosphate position of deoxynucleotides, through a P-N bond, metabolites such as taurine or aliphatic sulfonates are described and evaluated, both as pyrophosphate mimics and substrates for sulfonate delivery systems. Taurine-*N*-acetic acid was identified as the best leaving group in polymerase catalyzed DNA synthesis with various thermophilic and mesophylic DNA polymerases e.g. Therminator, Vent (exo-) and Klenow (exo-), and base specificity followed the order G > A ≥ T > C. Compared to the natural substrate dGTP, the V_{max}/K_M value for taurine-N-acetic acid-dGMP reduces by 702 fold, comparable to other bioisosteric leaving groups such as IDA and IDP.

In **Chapter 5**, the synthesis of various biotin-TMP and Kanamycin A-TMP conjugates, presenting some of the biocleavable linkers designed for PNCs, is described targeting biotin transport and aminoglycoside uptake systems, respectively. The biotin-TMP conjugates are linked through the 3'-OH group of TMP to the carboxyl group of biotin comprising an oxymethyleneoxy ester, a carboxy ester, a carbamate and an oxymethyleneoxyamide linker, whilst kanamycin A-TMP conjugates with carbamate and thioehtyl cabamate linkers were obtained by exploiting the different chemical reactivity of polyamimo and polyhydroxyl functionalities of the aminoglycosidic moiety.

In the last chapter (**Chapter 6**), the methods and results of biological evaluation of all conjugates of TMP with peptides, sulfonates, biotin and kanamycin using a thymidine auxotrophic *E.coli* mutant (thyA-, ΔphoA) is detailed. Unfortunately, none of those conjugates was found to be sustaining bacterial growth under the chosen experimental conditions.

Several critical parameters influencing the outcome of the biological assay can be identified, for instance conjugate concentration and mode of addition, stability, phosphatase activity, recognition and binding, metabolism, and need to be addressed in the future along with the exploration of additional linkers and conjugates.

Samenvatting

Het centrale dogma van moleculaire biologie steunt op de opslag, transfer en verbreiding van het genetische materiaal van DNA naar proteïnen via RNA. Synthetische genen kunnen worden gebruikt om genetisch gemodificeerde organismes te creëren. Deze organismes kunnen gebruikt worden voor verschillende toepassingen zoals therapeutics, landbouw, energie en milieu, materiaalkunde en biotechnologie. Maar dankzij het gevaar van al dan niet opzettelijk interacties van het gemodificeerd genetische materiaal met de genetische informatie van natuurlijke soorten via gene transfer, worden zulke artificiële genetische constructies gezien als een mogelijk gevaar voor natuurlijke ecosystemen.

De ontwikkeling van orthogonale biologische systemen door de chemische synthese van niet natuurlijke nucleïnezuren (Xeno Nucleic Acids, XNA), wordt beschouwd als een goede genetische firewall om genetisch cross talk te voorkomen. XNA lossen op deze mannier de voornaamste problemen in verband met bioveiligheid op. Verschillende op XNA gebaseerde systemen zijn gesynthetiseerd om deze firewall op te bouwen. Dit start met modificaties aan de base, het suiker en de leaving group, gevolgd door het aanpassen van de enzymes die geassocieerd worden met de XNA catalystise werking. Om hun ultiem evolutionair potentieel vrij te maken, moeten gereprogrameerde organismes een voldoende grote concentratie van voorlopers van XNA nucleotides of hun polymeren in de cel hebben. De *in vivo* invasie van XNA replicons was origineel gericht naar bacteriën vanwege hun relatief eenvoudige genetische architectuur. Hoewel de meeste gemodificeerde nucleosiden gemakkelijk bacteriecellen kunnen binnendringen met behulp van passieve diffusie, worden ze moeilijk gephosphorolyseerd tot hun monofosfaat vorm door natuurlijke kinases. Aan de andere kant wordt de levering van nucleotiden binnen de bacteriecel gehinderd door hydrofobe interacties met de lipidelaag van de cel. Dit komt doordat nucleotiden een lage pKa en een hoge polariteit hebben. Bacteriën hebben niettemin verschillende transmembraan transport proteïnen die door middel van actief transport essentiële voedingstoffen zoals peptiden, sulfonaten, suikers en vitamen binnen de bacteriecel kunnen transporteren.

Het doel van deze doctoraatsthesis is nagaan of deze natuurlijke systemen die gewijd zijn aan het transport van metabolieten, gebruikt kunnen worden als potentieel transportsysteem voor nucleotiden in bacteriecellen. Dit wordt onderzocht door het systematisch ontwerpen en synthetiseren van verschillende nucelotide-conjugaten, gevolgd door hun evaluatie in een voedingswaarde selectie assay. Thymidine mono-fosfaat (TMP) werd gekozen als voorbeeld nucleotide in een modelstudie waar de intracellulare levering van de corresponderende TMP-

conjugaten kon worden geanalyseerd door de groei van geschikte *E.coli* mutanten, die auxotrofisch zijn voor thymidine (thyA-, ΔphoA), te bestuderen.

Tijdens het eerste project (hoofdstuk twee) werd een varieteit van di- tot penta-peptides geselecteerd als transportmiddel, gericht naar verschillende peptide permeases, bijgevolg werd een serie van peptide-TMP conjugaten (PNCs) gesynthetiseerd. Deze werden geëvalueerd voor hun activiteit in *E.coli*. In dit hoofdstuk worden verschillende lineaire en convergente synthesemethoden in combinatie met een orthogonaal bescherming-ontscherming strategie besproken, om de conjugaten te synthetiseren. Deze conjugaten werden bekomen door de koppeling van een zijketen functionele groep van de oligopeptiden met de 3'-hydroxyl groep van TMP door middel van biologische kliefbare linkers zoals carbamaat (OCONH), ester (OCO), oxyamide (ONHCO), oxymethyleneoxyamide (OCH_2ONHCO) en oxymethylene-oxyester (OCH_2OCO).

In het derde hoofdstuk wordt het ontwerp en de synthese van verschillende PNCs of TMP bediscussieerd. Deze moleculen werden verbonden met behulp van verschillende fosfoester linkers zoals, fosfaat, fosforamidaat, fosforothioaat en thioethyl-fosfaat via de 5'hydroxyl van thymidine en de α-koolstof van de terminale C van glycine residue van L-Ala-Gly. Om een preliminair inzicht te krijgen in het potentieel van het *in vivo* metabolisme, werden deze conjugaten geëvalueerd als substraat in een templaat gerichte DNA polymerisatie reactie met verschillende thermofiele en mesofiele DNA polymerases zoals Therminator, Vent (exo-) en Klenow (exo-). Een vrije carboxyl groep is essentieel voor de binding van dit molecule in de actieve site van deze enzymes. De kinetische incorporatie data (V_{max}/K_m) van het conjugaat dat een fosforamidaatlinker en een vrije carboxyl groep bevat was 994 keer trager dan het natuurlijke substraat TTP. Dit resultaat is vergelijkbaar met eerder bekomen resultaten in ons labo.

In hoofdstuk vier werd de synthese van nieuwe sulfono-fosforamidaat conjugaten beschreven, waarbij de sulfono-fosforamidaat gekoppeld is aan de 5'-fosfaat postie van deoxynucleotiden door middel van een P-N binding. Metabolieten zoals taurine of alifatische sulfonaten zijn beschreven en geanalyseerd, zowel als pyrofosfaat mimics en als een substraat voor sulfonaat transportsystemen. Taurine-*N*-acetic acid was geïdentificeerd als de beste leaving group in een polymerase gecatalyseerde DNA synthese. Deze synthese werd uitgevoerd met verschillende termofiele en mesofiele Dna polymerases zoals Therminator, Vent (exo-) en Klenow (exo-). De base specifiteit had G > A ≥ T > C als volgorde. De V_{max}/K_m waarde voor taurine-*N*-acetic acid was 702 keer trager dan het natuurlijk substraat dGTP. Dit komt overeen met andere bioisosterische leaving groups zoals IDA en IDP.

In hoofdstuk vijf wordt de synthese van verschillende biotine-TMP en Kanamycin A-TMP conjugaten voorgesteld. Deze zijn respectievelijk gericht naar het biotine transport en het aminoglycoside opnamesysteem. De gesynthetiseerde conjugaten bevatten één van de biologische kliefbare linkers die ontwikkeld waren voor PNCs. De biotin-TMP conjugaten zijn verbonden via de 3'-OH group van TMP met de carboxyl groep van biotine, bestaand uit een oxymethyleneoxy ester, een carboxyl ester, een carbamaat of een oxymethylenoxyamide linker. De kanamycin A-TMP conjugaten bevatten een carbamaat en een thioethyl carbamaat linker. Deze werden bekomen door gebruik te maken van de verschillende reactiviteit van de polyamino en polyhydroxyl functionele groepen van het aminoglycosidic deel van het molecule.

In het laatste hoofdstuk (hoofdstuk zes) worden de verschillende methoden en de resultaten van de biologische evaluatie van alle TMP conjugaten met peptides, sulfonaten, biotin en kanamycin weergegeven, gebruik makend van het auxotrofisch *E. coli* mutant (thyA-, ΔphoA). Spijtig genoeg kon geen van de conjugaten bactieriegroei stimuleren, onder de gekozen experimenteële omstandigheden.

Verschillende cruciale parameters die het resultaat van de biologische assay kunnen beïnvloeden zijn geïdentificeerd: de conjugaat concentratie en de additie wijze, stabiliteit, phosphatase activiteit, herkenning en binding en metabolisme. Deze parameters moeten in de toekomst allemaal geanalyseerd en geverifieerd worden om mogelijke problemen uit te sluiten.

Abbreviations

1D	one dimensional
2D	two dimensional
A	adenine
ACN	Acetonitrile
L-Ala	L-alanine
L-Asp	L-aspartic acid
Bn	Benzyl
Boc	*tert* –butyloxycarbonyl
$(Boc)_2O$	di-*tert* –butyl dicarbonate
Bz	Benzoyl
C	cytosine
Cbz	Carboxybenzyl
COSY	Correlation SpectroscopY
DABCO	1,4-diazabicyclo[2.2.2]octane
dAMP	deoxyadenosine monophosphate
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
DMF	*N*,*N*-dimethylformamide
DMSO	dimethyl sulfoxide
DPDS	2,2'-dipyridyldisulfide
DPPA	diphenylphosphoryl azide
δ	chemical shift
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
D_2O	deuterium oxide (Heavy water)
ds-DNA/RNA	double-stranded DNA/RNA
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EtOAc	ethyl acetate
Gly	glycine
L-Glu	L-Glutamic acid
HETCOR	HETeronuclear CORrelation
His	histidine
HIV	Human immunodeficiency virus
HMBC	heteronuclear multiple bond correlation

HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence
Hz	hertz
IDA	iminodiacetic acid
L-Lys	L-Lysine
***i*-Pr**	isopropyl
***m*-CPBA**	*m*-chloroperoxybenzoic acid
Me	methyl
MHz	megahertz
MP	monophosphate
***n*-Bu**	*n*-butyl
NDP	nucleotide diphosphate
NHS	*N*-hydroxy succinimide
NMI	*N*-methylimidazole
NMP	nucleoside monophosphate
NMR	nuclear magnetic resonance
***n*-Pr**	*n*-propyl
NPs	Nucleoside 5′-phosphoramidates
NTP	nucleoside triphosphate
MMTr	mono methoxy (4-) trityl chloride
PCR	Polymerase chain reaction
Ph	phenyl
pTSA	para-toluene sulfonyl chloride
L-Phe	L-Phenyl alanine
Ppm	parts per million
Py	pyridine
RNA	ribonucleic acid
RT	reverse transcriptase
rt	room temperature
SATE	*S*-acylthioethyl
T	thymine
TBDMSCl	*tert*-butyldimethylsilyl chloride
TBDPSCl	*tert*-butyldiphenylsilyl chloride
TIBSCl	2,4,6-triisopropylbenzenesulfonyl chloride
TEA	triethylamine
TEAB	triethylammonium bicarbonate

TFA	trifluoroacetic acid
THF	tetrahydrofuran
TK	thymidine kinase
TLC	thin layer chromatography
TMSCl	Chlorotrimethylsilane
(d)TMP	Thymidine monophosphate
TOCSY	TOtal Correlation SpectroscopY
TTP	Thymidine triphosphate
U	uracil

TABLE OF CONTENTS

Chapter I. General Introduction

1.1. Nucleic acids: A brief history

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are the two essential biopolymers responsible for the storing and transmitting genetic information, from generation to generation in all living organisms. The Swiss scientist Friedrich Miescher was the first discoverer of DNA back in 1869, but only in 1944 Avery and co-workers identified the important biological function of DNA as genetic material. Generally, in an eukaryotic cell, DNA is present inside the nucleus as a highly packed coiled structure, as complex with histone proteins. DNA is predominantly found in nature in a double-stranded anti-parallel helical form, while RNA is largely single-stranded. Francis Crick and James D. Watson first presented the double helical molecular structure of DNA in 1953[1] and they were awarded Nobel Prize together with Maurice Wilkins in 1962. Their findings ultimately led to the opening of the new field known as 'Chemical Biology', and building upon this idea scientist later discovered how DNA function as the storage of the genetic information, which then transfer the genetic information to RNA. RNA functions as an intermediate to transfer the genetic information from DNA to the protein level via a three-letter genetic code. Such information flow was first proposed (1956) by Francis Crick as the "central dogma" of molecular biology (Fig. 1-1), "*DNA makes RNA makes protein*".

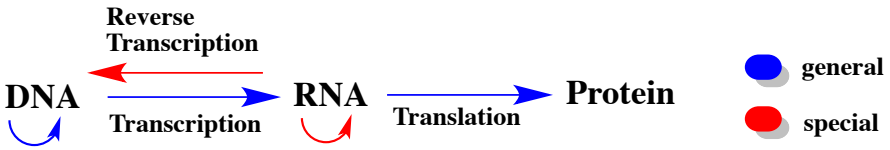


Figure 1-1. Information flow in biological systems.

1.1.1. Chemical structures of nucleosides and nucleotides

DNA and RNA are both constructed from the four different nucleotide monomeric units by the repetition of phosphodiester linkages. Each nucleotide unit is comprised of a nucleobase (a heterocyclic nitrogenous base), a five-carbon sugar (either deoxy-ribose or ribose) and at least one phosphate group (Figure 1-2). Without the phosphate group, the nucleobase and sugar moiety together are denoted as a nucleoside and with respect to the attached phosphate group/s, the corresponding nucleotide can also be called a nucleoside monophosphate, diphosphate and triphosphate respectively. In the case of DNA, the sugar moiety is 2′-deoxy-D-ribose, which is connected to a nucleobase *via* a *β*-glycosidic C-N bond at the 1′-carbon position. Nucleobases are divided in two families, purines [adenine (A) or guanine (G)] or pyrimidines [cytosine (C) or thymine (T)]. The sugar found in RNA instead is D-ribose, and the same nucleobases can be attached to the ribose moiety as found in DNA,

with the exception of thymine, which is replaced by uracil (U). The nucleoside triphosphates [d(NTPs)] of the corresponding [deoxy-(nucleoside monophosphates)] act as substrates for polymerases and pyrophosphate as the metabolic active function necessary for nucleic acid biosynthesis *in vivo*.

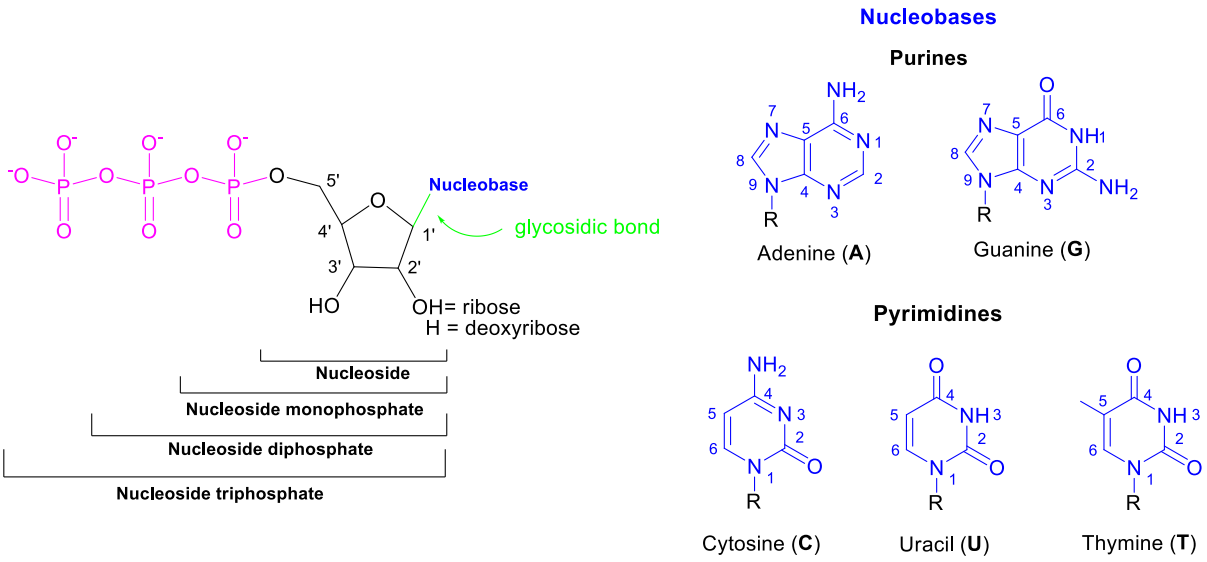


Figure 1-2. Chemical structures of natural nucleotides (left), and the five heterocyclic nucleobases (right) with the conventional atom numbering.

1.1.2. Secondary structures of DNA and RNA

Nucleic acids are built up via the formation of a repeating phosphodiester bond following the nucleophilic attack of the 3'-OH group of one nucleotide to the 5'-*O*-phosphate of another nucleotide with the aid of a suitable enzyme. Mostly, DNA occurs in nature as a double stranded helix, for at least some of their lifespans; whereas RNA as single stranded. The Watson-Crick type base pairing of the two strands fundamentally determines the secondary structure of DNA, where adenine is paired with thymine through two hydrogen bonds, and guanine is paired with cytosine through three hydrogen bonds (Fig. 1-4). Different conformations of a particular nucleotide, as represented by the backbone and glycosidic torsional angles, and the puckering mode of the sugar residues lead to different overall helical duplex conformations.

The furanose sugar ring of DNA/RNA adopts a non-planar conformation that can be described as a twist of the C2'-C3' bond with respect to the plane defined by C1'-O4'-C4' atoms. When the 5'- and 2'-carbon share the same plane in the puckering orientation the resulting conformation is called C2'-*endo*, whilst when the 5'- and 3'-carbon share the same plane in the puckering orientation is called C3'-*endo* (Fig. 1-5).[2] Generally DNA-sugar adopts a South type puckering conformation,

while RNA-sugar prefers a north type puckering (Figure 1-5). Depending on the sequence context, ionic and hydration situations, the DNA-sugars can favor any of these puckering modes alone or in combination and thereby considerably influence the three-dimensional structure of DNA double helix.[3]

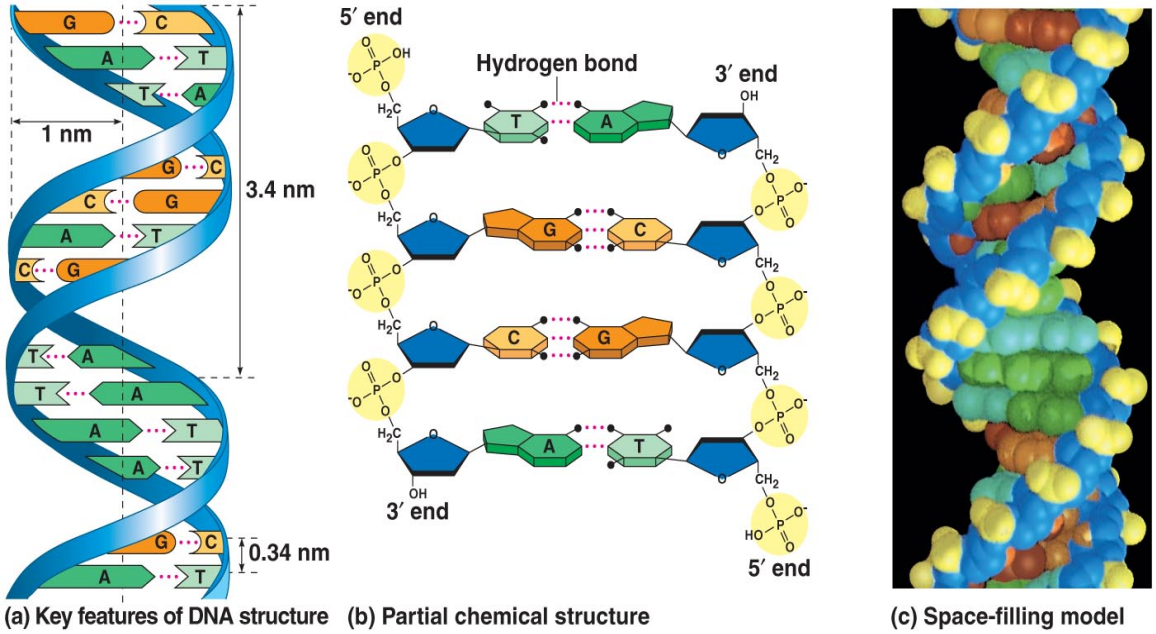


Figure 1-3. Three-dimensional structure of a DNA-duplex (reproduced from http://academic.brooklyn.cuny .edu/biology/bio4f v/page/molecular%20biology/dna-structure.html)

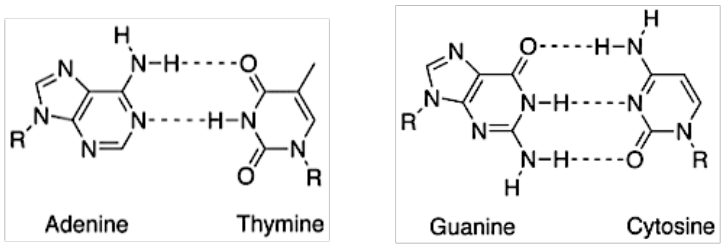


Figure 1-4. Classical Watson-Crick base pairing in DNA

The nucleobases can also adopt two different orientations, *syn*- and *anti*- with respect to the sugar unit by rotation along the glycosidic bond. In the *anti*-conformation the bulk of base is away from the sugar ring and in the *syn*-conformation the bulk is closer to the sugar ring. Due to the steric crowd, generally the *anti*-conformation is more favored and abundant than the *syn*-form.

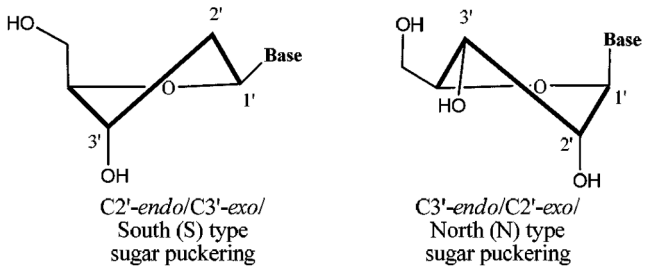


Figure 1-5. The different arrays of non-planar sugar conformation of the pentofuranosyl sugar ring of DNA (left) and RNA (right).

1.2. Synthetic biology

Synthetic biology is an emerging, interdisciplinary area of research in the field of biology, which bring together disciplines such as genetic engineering, molecular biology, systems biology, biotechnology, biophysics, and also evolutionary biology. The term "synthetic biology" dates back as early as the twentieth century, but it is only during 1970 that the polish geneticist Wacław Szybalski used it in a modern sense, writing: "*Let me now comment on the question 'what next'. Up to now we are working on the descriptive phase of molecular biology…. But the real challenge will start when we enter the synthetic phase of research in our field. We will then devise new control elements and add these new modules to the existing genomes or build up wholly new genomes. This would be a field with an unlimited expansion potential and hardly any limitations to building 'new better control circuits' or ..... Finally other 'synthetic' organisms, like a 'new better mouse'. ... I am not concerned that we will run out of exciting and novel ideas, ... in the synthetic biology, in general.*"[4]

In 1978 Arber, Nathans and Smith won the Nobel Prize in Physiology and Medicine for the discovery of restriction enzymes, their work eventually leading to a boom in the construction of recombinant DNA molecules and analysis of individual genes. This represented a milestone in the field as not only existing genes could be defined and investigated but also new gene arrangements could be fabricated and evaluated. Although this field has been growing very rapidly over the last 15 years, a broadly acknowledged single definition does not exist yet, but the main objective is "manipulating and fabricating biological devices, systems or pathways for our beneficial purposes, which do not exist in nature."

1.2.1. Applications of synthetic biology

Synthetic biology is now no longer constrained to an abstract concept, but many challenges have started to be addressed and some significant progresses have been achieved by producing novel chemical entities, addressing environmental issues and thus improving overall human health. Probably, the first successful applicative example is the development of recombinant DNA technology which allowed to engineer *E.coli* bacteria for the large scale production of the human growth hormone to help under-growing individuals[5] and human insulin for diabetic

patients.[6] A similar principle was applied to engineer microbes for the production of most of the vitamins that are being consumed as supplements.

Outstanding progress has been made in the cost effective production of synthetic biofuels from different cheap sources to address the global warming and energy self-reliance issues.[7] Different groups around the world also achieved ground breaking discoveries on the field of 'biosensors' by using engineered bioluminescent bacteria that can detect different environmental pollutant like mercury, arsenic, petroleum pollutant through quorum sensing.[8] As the world population is growing exponentially and so does the expected acute food shortage to feed the hungry world, remarkable efforts have been made through bio-engineered crops (green revolution) by joint efforts of different agro-biotech companies and academics from different universities. In the modern time the pharmaceutical industry has engineered proteins to be used as monoclonal antibodies and drug target mimics, required for high throughput screening. To tackle complex diseases like cancer or auto-immune diseases, targeted complex engineered cells or proteins are required for screening purposes.[9] Besides, synthetic biology was successfully employed in material science for the production of reengineered curli nanofibers for substrate adhesion, protein immobilization and nanoparticle templating.[10] Along with the significant number of examples and on-going research projects, huge are the opportunities left for successful applications of synthetic biology that can be envisioned. With the advancements in this field, expectations are also increasing to engineer organisms with genuinely new features from what is already known, for example not just confining into the familiar recombinant technology, but also achieving the true evolution of synthetic organisms with alternative genetic systems.

1.2.2. Synthetic biology and biosafety

As a consequence of the incorporation of synthetic DNA constructs together with synthetic enzymes in chromosomes and episomes of modified microorganisms for applications in agriculture, energetics and molecular biology, fundamental ethical questions and biosecurity issues due to the potential contamination of synthetic genetic sequence with natural system are raised. These artificial constructs are considered by some as a threat to the wild ecosystem and even alleged to manipulate the evolution of terrestrial life. These issues are not new, but were addressed by the regulation of companies producing synthetic DNA or enzymes,[11] during the early stages of development of the recombinant DNA technology and launching of genetically modified organisms (GMOs) in agriculture to improve crops yield and disease or insect resistant transgenic plants.[12] There are, however, also possible and perceived genetic pollution due to deliberate or unintended damage. There are already strict regulatory board for extensive guidelines for genetic engineering and synthetic biology in all over the world. SYNBIOSAFE and COSY are EU funded organizations which function in Europe to care for the safety related issues regarding synthetic biology.

1.3. Xenobiology: The crucial biosafety tool

"It is when we all play safe that we create a world of utmost insecurity."- Dag Hammerskjöld.

To address the issue of biosafety, an important component in the development of technologies that must be emphasized is the avoidance of crosstalk between natural species and artificial biodiversity. Standardized DNA parts (BIOBRICKS), cost effective DNA synthesis, reliable DNA sequencing and molecular modeling prior to fabrication, are the key enabling technologies for advancement in this field. As it is now becoming easier to construct artificial genes, plasmids and chromosomes using standard protocols with the help of these technologies, the moon-shot goal of genetic reprogramming, directed mutagenesis and directed evolution could be anticipated. As an effort towards suitable biological systems devoid of cross contamination, it was proposed to design an orthogonal episome, i.e. a self-sustaining third type of information system different from DNA or RNA, which would not interfere (no vertical or horizontal gene transfer) with natural ecological system.

1.3.1. Xeno nucleic acids (XNAs)

In order to create a parallel system in any discipline, it is necessary to start from the elementary building blocks. As the nucleic acids are the bio-systems carrying the natural genetic information, so to avoid cross talking interference, substitution of all of their fundamental units must be substituted by modified or 'xeno' nucleic acids. Enormous efforts have been made to produce nucleic acids modifications, not only in this field but also in the medical one, for potential use as drugs especially in antiviral therapy, anti-gene, antisense and siRNA and also in other area like cellular probes, biosensors, aptamers and enzymes in biotechnology. Over the past few decades, chemistry based research has clearly established a third type of molecular system for storing genetic information other than DNA and RNA.[13] XNAs can different from natural nucleic acids from various aspects, as chemical modifications can be tailored either at the hetero-cyclic base moiety or at the sugar moiety or phosphodiester linkage or a combination. Recent progresses in the field of engineered polymerases to achieve enzymatic replication of unnatural nucleic acids paved the path to the synthesis and evolution of unique synthetic polymers with the desired properties and functions. Moreover, the base pairing ability of XNA with the desired oligonucleotide polymers, template based enzymatic incorporation fidelity of XNA, transcription and reverse-transcription of genetic message via XNA gene, are the key principles based on which natural nucleic acid mimics are recognized for the application in the synthetic biology field.

1.3.2. Base modifications

There are numerous examples listed in the literature of modified nucleo-bases. While some of them have been designed for genetic and synthetic biology purposes (Fig. 1-6), others derive from therapeutic purposes. Different groups focused on the development of non-natural base-pairing

systems for incorporation into DNA and RNA, and in turn to expand the genetic alphabet. There are two main strategies to expand the genetic code, a) maintaining the Watson-Crick like hydrogen bonding mode in the modified base pairs, b) using hydrophobic groups that mimic the shape and polarity of natural bases through hydrophobic interactions (Fig. 1-6-1). Besides these, other strategies include size expansion, extra H-bond and metal mediated pairing (Fig. 1-6-1). In late 1980, the Benner group first introduced the several unnatural base pairs showing a Watson-Crick type hydrogen bonding pattern, such as isoguanine:isocytosine (isoC:isoG) (Fig. 1-6-2B), 6-amino-5-nitro-3-(1′-β-D-2′-deoxyribofuranosyl)-2(1H)-pyridone:2-amino-8-(1′-β-D-2′-deoxyribofuranosyl)-imidazo[1,2-*a*]-1,3,5-triazin-4(8H)-one (Z:P) (Fig. 1-6-2B) and 5-(2,4-diaminopyrimidine:xanthosine (dK:dX) for synthetic biology purposes based on AEGIS (artificially expanded genetic information systems).[14] The authors first successfully explored the isoC:isoG pair for its ability to be incorporated and replicated by the Klenow fragment of *E.coli* DNA polymerase. However, to maintain high isoC:isoG selectivity the simultaneous addition of 2-thiothymine (2-thioT) (Fig. 1-6-2A) was crucial to reduce isoG:dT base pairing, by forming strong 2-thioT:A pairing system.[15] They have also described that high fidelity and selectivity (97.5% retention) can be maintained with Z:P base pairs[16] both in DNA polymerization and PCR amplification.[17]

Eric T Kool and co-workers first introduced the concept of xDNA with 8-component bases where the modifications were made for four natural nucleobases A, T, G, C to increase the molecular volume by introducing an extra phenyl ring in the bases: xA, xT, xG, xC or yA, yT, yG, yC (Fig. 1-6-2A).[18] [19] The modified bases could also pair with the corresponding natural counterparts, and in a recent structural study it was shown that the greater the distance between the wider grooves and longer pitch of the double helix were responsible for the more stability of xDNA compared to natural DNA.[20] The same group also presented an analogous yDNA using the size expansion concept with a phenyl ring. Recently, the first living cell study in *E.coli* revealed that the two dXNA bases xA and Xc were recognized and replicated efficiently by *E.coli* polymerase to develop a normal message.[21]

Apart from the hydrogen bonding modulation, efforts have also been made to mimic size, shape and polarity of a natural pairing system by a hydrophobic approach. Furthermore, Kool et al. developed an hydrophobic unnatural pair (Z:F) (Fig. 1-6-1), mimicking the canonical A:T base pair and found that the difluorotoluene which mimics thymine was also incorporated and replicated by DNA polymerase.

Since Kool's discovery of hydrophobic interactions for base modification, Hirao and co-workers produced the 7-(2-thienyl)-imidazo[4,5-b]pyridine (Ds):pyrrole-2-carbaldehyde (Pa) (Fig. 1-6-2C) and Ds:2-nitropyrrole (Pn) base pairs aiming at the expansion of the genetic alphabet, which retained high selectivity and fidelity per cycle (>99%) in PCR in the presence of a DNA polymerase with

3′→5′ exonuclease activity.[22] The simultaneous addition of γ-amido triphosphates was necessary to maintain high selectivity, but resulted in reduced fidelity. Building up on this idea to increase fidelity in PCR, they also developed a similar Ds: 2-nitro-4-propynylpyrrole (Px) base pair, with the aim of generating several functional assemblies of interest at the propargylic position of Px.[23] Site specific incorporations with high selectivity were observed for the Ds:Px pairing system, in a PCR amplification of 10^7-fold by 30 cycles and with more than 99.9% fidelity per cycle, without the need of external addition of γ-amido triphosphates.

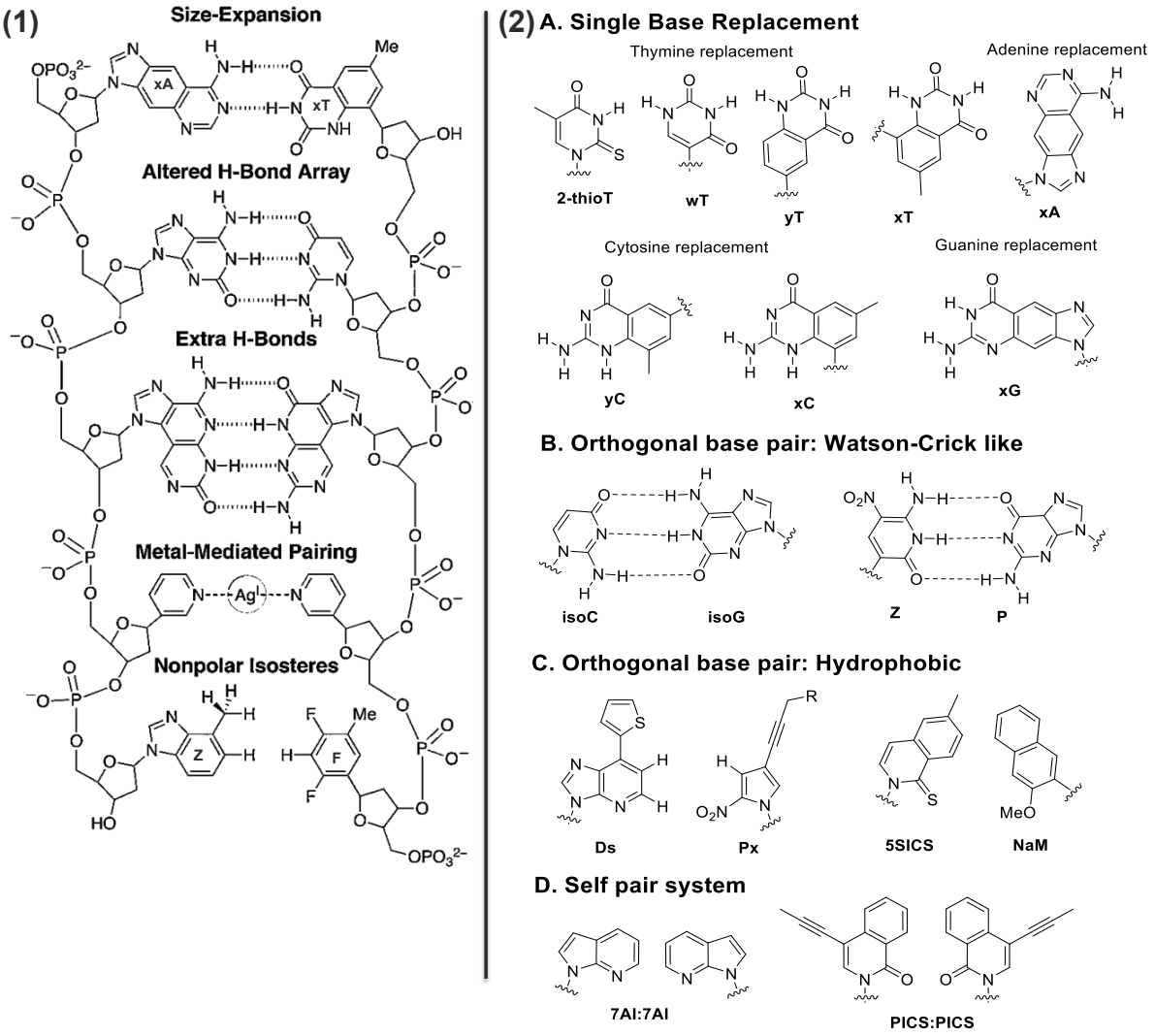


Figure 1-6. Examples of unnatural base pairs with different types of bonding systems

Romesberg's group also independently studied the modification of hydrophobic base pairs in a systematic way aiming at the extension of genetic alphabet. Various hydrophobic bases were synthesized and screened for their pairing properties like d5SICS:dMMO2 (Fig. 1-6-2C), 7AI:7AI, PICS:PICS (Fig. 1-6-2D), d5SICS:dNaM and dTPT3:dNaM.[24] Notably, the d5SICS:dNaM pair has drawn lot of attention since it could be both replicated and transcribed *in vitro* as well as *in vivo*.[25, 26]

1.3.3. Sugar modifications

A large number of studies have been devoted by different laboratories to develop xeno-nucleic acid systems by introducing sugar modifications mimicking the canonical ribose or deoxy-ribose sugar. Not only the sugar modification has potential applications in synthetic biology (summarized in fig. 1-7), but also in therapeutics, especially in antiviral therapy.

Locked nucleic acids often abbreviated as LNA, were introduced independently by two groups, Imanishi and Wengel almost 15 years ago. Since then many laboratories have been working on those sugar modified nucleic acids for their potential applications as antisense oligonucleotides, molecular beacons, DNAzymes etc.[27] LNAs mimic the corresponding natural RNA moiety, where the 2'-O and C-4' are bridged with an extra methylene group. This locking system allows the sugar ring to adopt a stable 3'-*endo* conformation which reproduces the natural RNA binding. Due to the reduction in conformation flexibility, LNA has an exceptional binding affinity for self-pairing and also towards natural DNA and RNA counterparts. Structural NMR spectroscopy studies revealed that a LNA-RNA duplex adopt an A-type double helical structure where a LNA-DNA duplex has a structural similarity to those of a natural RNA-DNA duplex.[28] The incorporation properties of LNA were studied using different DNA polymerases like Vent(exo-) and KOD(exo-) and also with T7 RNA polymerases, giving encouraging results.[29] Other variants of locked nucleic acid like *xylo*-LNA, 2'-thio LNA and 2'-amino LNA are also being explored for various applications.

Amongst the various nucleic acid mimetics developed by Eschenmoser et al., L-α-threofuranosyl nucleic acid (TNA) is of particular relevant to synthetic biology due to the inherent stability of this polymeric unit in the presence of nucleases.[30] Unlike DNA, phosphodiester bonds in TNA occur between the 2'- and 3'- position of the furanose ring (3'→2') and no methylene group is present between the sugar rings. TNA polymers were found to be capable of exchanging information by forming intra-system duplexes with complementary TNA templates and also cross pairing with DNA and even more strongly with RNA. Structural studies by X-Ray crystallography revealed that a TNA-TNA duplex prefers to adopt a structural similarity to the A-form of DNA or RNA, but with a relatively shorter distance between the repetitive units as anticipated.[31] Due to its structural simplicity and binding affinity, scientists have hypothesized that TNA could have served as an ancestral genetic system alternative to RNA in the prebiotic pool. The replication properties of TNA have been studied independently by Szostak and Herdewijn using different DNA polymerases and reverse transcriptases, but reported to be kinetically less efficient than natural analogues. Recently, TNA triphosphates were efficiently incorporated by an engineered reverse transcriptase (RT521) on a DNA template, and TNA sequences were generated up to 72 nt following canonical base-pairing selectivity and transcribed into DNA up to 93 nt based on a TNA template using the same polymerase.[32]

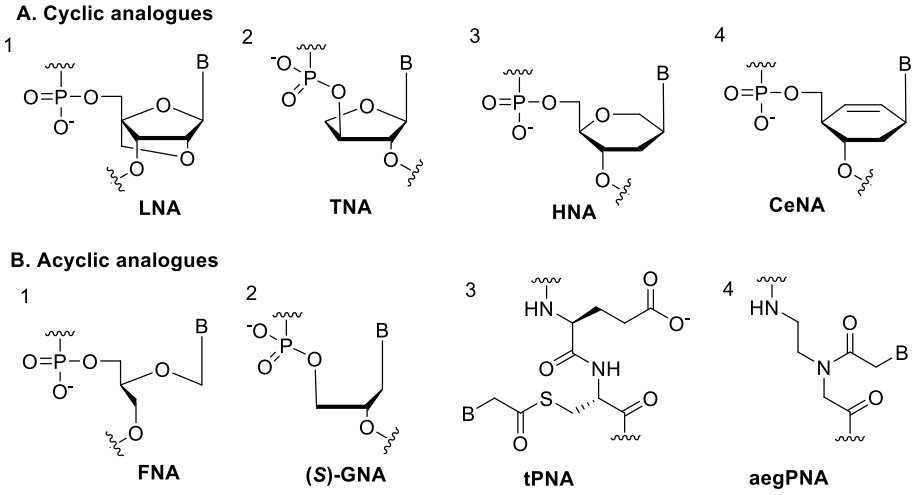


Figure 1-7. Examples of sugar and backbone modified nucleic acid analogues.

Among the six-membered sugar mimics, Hexitol Nucleic Acid (HNA) and Cyclohexenyl Nucleic Acid (CeNA) are the two promising candidates introduced by Herdewijn and co-workers in the late 1990s. HNA is similar to DNA with a phosphorylated 1', 5'-anhydrohexitol backbone bearing an extra methylene group present between C-1' and O-4' of the ß-D-2'-deoxyribose unit. HNA showed inter-system cross pairing with DNA, and its antisense properties are known due to its stronger binding with RNA,[33] furthermore HNA homo-duplexes are even more stable due to stronger intra-system cross pairing. Like LNA, it also forms A-type helices of HNA-HNA and HNA-RNA duplexes.[34] The other six-membered sugar mimic CeNA is characterized by a flexible cyclohexene ring with a ß-positioned nucleobase like in HNA. Due to the intra- and inter-system pairings and its prolonged stability in serum, CeNA is also considered as a potential antisense candidate.[35] Both the triphosphates of HNA and CeNA showed moderate to good incorporation properties in the presence of different DNA-polymerases. Deep Vent (exo-) showed good incorporation efficiency and kinetics of incorporation were comparable to the natural analogues but with reduced fidelity. Cyclohexenyl nucleotides also showed good incorporation properties in the presence of Vent(exo-) DNA polymerase and HIV-reverse transcriptase, showing up to seven consecutive strand elongation based on a DNA template containing T and C.[36] Recently both HNA and CeNA triphosphates were efficiently incorporated up to 72 nt based on a DNA template by engineered DNA polymerase PolC7 and Pol6G12 respectively. Similarly, DNA sequences were also reverse-transcribed up to 72 nt on CeNA template and 73 nt on a HNA template using engineered reverse transcriptases RT521K and RT521 respectively.[32]

For their structural simplicity and conformational flexibility, acyclic nucleic acids were also explored early in the origin of life context, where they might have existed as precursors for non-enzymatic oligomerization in the pre-RNA world.[37] [38] Some of these flexible nucleic acids still maintain canonical base pairing which is essential for polymerase-catalyzed incorporation studies.

FNA (flexible nucleic acid), based on formyl glycerol, is one of the initial acyclic nucleic acids that were thought to dominate in the prebiotic world, as hypothesized by Orgel.[37] Though FNA showed weak or no tendency to form stable duplexes with DNA, it was lately found by Switzer et al. that both the enantiomers of FNA can be incorporated by various DNA polymerases and with the higher efficiency for the (R)-isomer. Therminator DNA polymerase showed the highest fidelity, with incorporation up to 7 nt.[39]

Like FNA, glycerol nucleic acid (GNA) is also highly flexible due to its propylene glycol phosphodiester backbone, although only particularly (S)-GNA has cross-pairing properties with RNA, whilst both the isomeric forms give relatively more stable self-paired homoduplexes.[40] Szostak was the first to examine the ability of GNA-triphosphates for their ability to be incorporated on a DNA template by different DNA polymerases, with therminator polymerase giving the best results with incorporation of 5 GNA unit.[41] Although GNA does not make stable strands with DNA, recently the formation of a full length DNA chain was established on a BstDNA polymerase catalyzed incorporation study based on a GNA template as reported by the same group.[42]

One of the most promising examples of nucleic acid mimics is represented by peptide nucleic acid (PNA). Aminoethylglycine PNA (aegPNA) was first developed by Nielsen and Buchardt as a ligand to bind to double stranded DNA via Hoogsteen base pairing.[43] In aegPNA, the original sugar-phosphate backbones were replaced by flexible achiral *N*-(2-aminoethyl) glycine pseudo-peptide polymeric units linked to nucleobases. Because of the formation of both homo-duplexes and hetero-duplexes with DNA and RNA by Watson-Crick pairing with high thermal stability and sequence specificity, PNA has attracted researchers for its diverse applications in biomedical research, i.e. genome mapping, mutation detection, anti-sense and anti-gene therapy etc.[44] The enzymatic incorporation of PNA is not known, but non-enzymatic polymerization of PNA based on DNA or RNA or the vice-versa using pre-activated monomeric units, has been reported by Nielsen and Orgel[45] PNA has also been studied in the context of the pre-RNA world because of its simplicity, structural rigidity, robustness and probable spontaneous polymerization at the boiling point of water.[46] In the last few years lot of research efforts have been made to introduce other variants of PNA like thioester PNA (tPNA) and also to enhance the bioavailability and to increase the potency of PNA. tPNA[47] is constructed from a alternating sequence of cysteine and non-cysteine residues, preferably amino acids with polar, ionic side chains along a polypeptide backbone. The thioester linkage was chosen in view of the reversible tethering between nucleobase thioester monomers and cysteine residues in solution, via a chemoselective transesterification with fast reaction kinetics. Upon incubation, the tPNA enriched with the sequence of nucleobases complementary to the oligonucleotidic template is formed and this unique self-assembly process could shed light on replication mechanism in a synthetic biology construct. Although the choice of PNA is not a very popular one in synthetic biology, as a result of the higher fidelity and compatibility in the template based replication, PNA could be an important candidate.

1.3.4. Leaving group modifications

In order to design XNAs and corresponding artificial genes able to propagate *in vivo*, multiple chemical and biological properties for each structural component have to be kept in mind. For instance, nucleoside triphosphates (NTPs), endorsed by high energy phosphono-anhydride bonds, possess multiple biological roles such as information storage and transmission (ATP as a building block for RNA biosynthesis), energy transfer (ATP as an energy transporter), regulatory tuning (at the level of kinases and signal transduction) and proliferating metabolic corridors (ATP as an activated carrier molecule). So it was thought due to its multifunctional role, the triphosphate group might be a limiting factor for XNA-propagation *in vivo* using synthetic XNA-triphosphates. Additionally it was also proposed that the use of alternative leaving groups (ALGs) mimicking the function of pyrophosphate in NTPs could serve the purpose of an information system *in vivo* simultaneously eluding interference of other cellular function.[48] Ideally, XNA-based engineered polymerases could only recognize the modified nucleotides as building blocks for gene synthesis, not accepting natural NTPs as substrates and vice versa, ALGs should not be accepted by any natural enzymatic machinery.

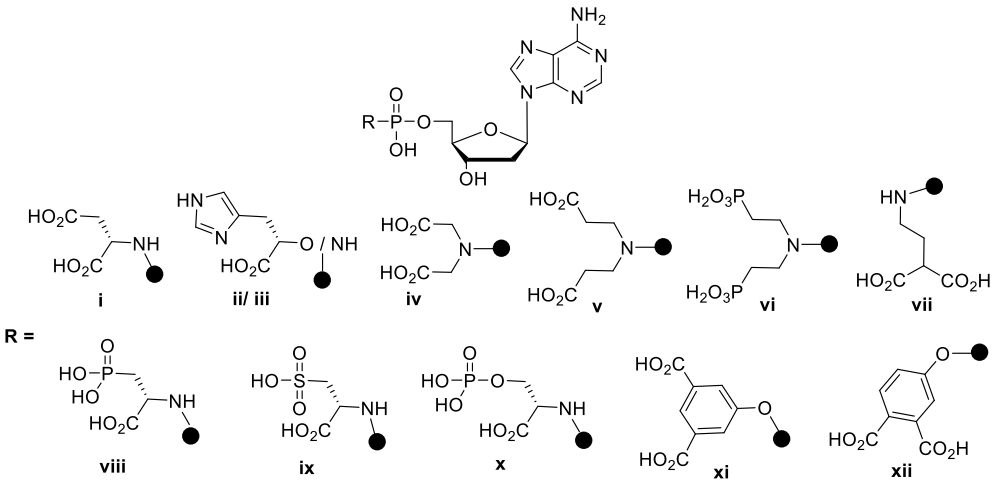


Figure 1-8: Chemical structures of modified nucleotide analogues with alternative leaving groups mimicking the pyrophosphate group.

Prior to Herdewijn's work, minor modifications like methylene phosphonates, phosphoramidates and halo-triphosphates were reported in the literature as triphosphate mimics, generally as substrate for HIV and AMV reverse transcriptases.[49] Herdewijn and co-workers made important contribution in this area by developing ALG-system which could mimic pyrophosphate in the enzyme-catalysed nucleic acid synthesis. As the nucleotide conjugates comprising ALGs do not occur in cellular media, to achieve *in vivo* XNA polymerisation, such synthetic precursors need to be supplied from outside the cell. For their successful application many prerequisites have been defined for substrates with an ALG, listed as follows:

- Good water solubility and optimal chemical and enzymatic stability in extracellular media

- Should fit in the active site of the enzyme and exhibit good substrate property
- The polymerase should mechanistically be able to release the leaving group
- To ensure the irreversible polymerization process, any leaving group should be actively degraded or reused as common non-toxic cellular metabolites
- Ability to be taken up actively inside the cell.

Among the diverse range of leaving group modifications that were examined, some of the best illustrative examples are presented in Fig 1-8. The first ALGs that were recognized by polymerases were found to be two amino acid L-Asp and L-His, amongst the different amino acids screened. Whilst, L-Asp-dAMP and L-His-dAMP showed good substrate property mimicking natural dATP in DNA chain elongation studies catalyzed by HIV RT and Therminator DNA polymerase.[50] In case of HIV-RT, the efficiency of L-Asp-dAMP was higher than that of L-His-dAMP, but the fidelity of DNA synthesis slowed down after 2 nt. The L-Asp variant with all the nucleotides (L-Asp-dAMP, L-Asp-dTMP, L-Asp-dGMP and L-Asp-dCMP) were also accepted for DNA incorporation by HIV-RT maintaining Watson-Crick base pairing.[51] From the steady state kinetics it was observed that despite the maximum initial rate (V_{max}) of L-Asp-dAMP incorporation was comparable with the natural analogue dATP, the relatively high K_m value of L-Asp-dAMP compared to dATP implies weak binding to the active site. Following these initial encouraging results, different leaving groups with phosphodiester linkages were also examined to compare the activity with phosphoramidate linkage containing ALGs. For example, the maximum velocity of ß-imidazole lactic acid-dAMP (ILA-dAMP) was higher by ten-fold than L-His-dAMP, while substrate affinity by two-fold.[52] [53] Likewise, nucleotide analogues of iminodiacetic acid (IDA), iminidipropionic acid (IDP) and 3-phosphono-L-alanine were synthesized and evaluated as potential substrates for HIV-RT.[54]

IDA and 3-phosphono-L-alanine conjugates with all four bases were synthesized to study the influence of the base moiety, and it was found that the incorporation efficiency decreases in the order A > T > G > C. Furthermore, the steady state kinetic analysis of phosphoramidate analogues of adenine IDA-dAMP, IDP-dAMP and 3-phosphono-L-Ala-dAMP indicated that catalytic efficiency (V_{max}/ K_m) with respect to the natural substrate dATP, is significantly declined by 940, 83 and 99 fold respectively. The lower specificity of the modified substrate may be due to lower binding affinity to the active site of commercially available polymerases.

Along with the systematic investigation of the ALGs retaining the canonical base pairing, the directed XNA based evolution in the active site of nucleic acid polymerases was also explored for enhanced selectivity. The efficacy and fidelity of polymerase catalyzed synthesis of modified nucleic acid polymers from unnatural nucleoside triphosphates are generally low, due to proper selectivity of the polymerase towards the corresponding XNA precursor. With the advent of directed polymerase evolution techniques (like phage display[55], CSR[56] and CST[32]), it now became relatively easy for fabricating engineered polymerases which can efficiently incorporate XNA precursors during the polymerization process. Considering the different enzymatic actions like copying, transcription,

translation and reverse-transcription, different DNA/RNA dependent XNA polymerases and XNA dependent DNA, RNA and XNA polymerases were also envisioned for the successful *in vivo* propagation and evolution of XNA.[57] Recently, Holliger and co-workers made fundamental progress in enzyme evolution by engineering polymerase mutants able to support *in vitro* assembling and replication of genetic messages for six types of sugar-modified XNAs (HNA, CeNA, ANA, FANA, TNA, and LNA). These engineered polymerase mutants are able to recognize and efficiently incorporate XNA building blocks from a DNA template. XNA sequences were generated up to 72 nt following canonical base-pairing selectivity and reverse-transcribed into DNA with 95% accuracy or more.[32]

1.4. Orthogonal information system and importance of nucleotides uptake in bacteria

To address the biosafety issues linked to synthetic biology, the construction of a third type of 'orthogonal' nucleic acid system (episome) was postulated by Herdewijn and Marliere, with the foremost aim of preventing genetic cross talk with existing ecosystem, by acting as a genetic firewall.[57] In mathematics the term 'orthogonal' indicates two vectors that are perpendicular to each other or generally, things, which are non-related to each other. Applied to synthetic biology this concept implies *in vivo* propagation of modified nucleic acids (XNAs), which structurally differs from natural DNA or RNA, but can autonomously replicate and perform the necessary biological functions not interfering with natural biological pathways. For the long term goal of solving the 'orthogonal episome' puzzle, several significant correlated footsteps need to be examined independently and methodically which includes base, sugar and alternative leaving group modifications together with directed XNA based enzymes engineering. The proliferation of XNAs along with their well-tailored XNA based polymerases and XNA-zymes in an *in vivo* scenario is a truly multidisciplinary effort comprising chemistry and various fields of biology.

Concerning the nucleobases, various single modifications have been carried out to mimic the natural canonical base pairing systems e.g. dP:dZ, isoC:isoG, ImN^{O}:NaO^{N}, Ps:Dx and d5SICS:dNaM. Several alternative leaving groups e.g. L-Asp, IDA, IDP and 3-phosphono-L-Ala were examined which exhibit potential pyrophosphate mimicking properties. Attempts towards the backbone modification e.g. HNA, CeNA, LNA, TNA have also been introduced different from DNA or RNA. It was also demonstrated by Holliger group, by fabricating engineered polymerases it was also possible *in vitro* efficient DNA based XNA incorporation and XNA based DNA reverse-transcription.[32] Recently, with the help of XNA replication technology, the same group has developed various XNAzymes based on four XNAs ANA, FANA, HNA and CeNA, that are able to mimic natural catalytic activities e.g. RNA endonuclease, RNA ligase and also XNA-XNA ligase activity on a FANA network. [58]

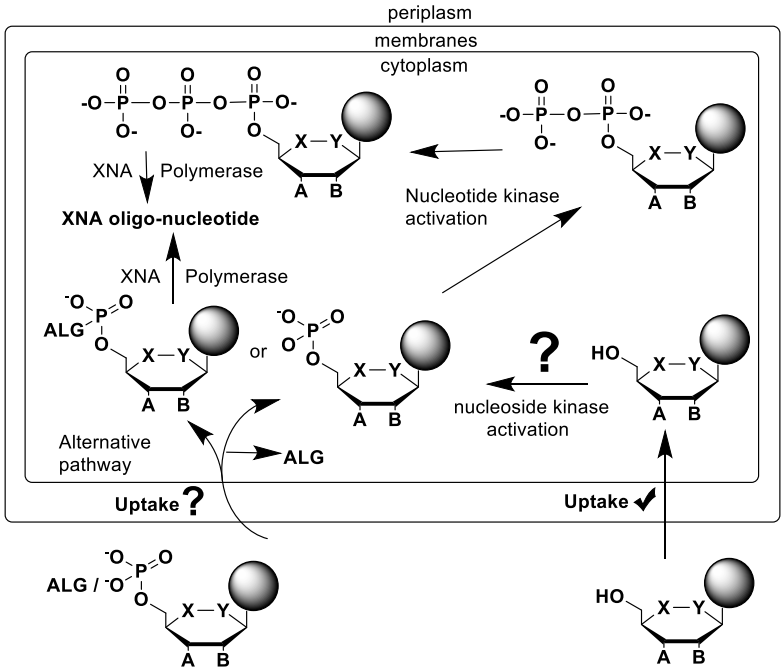


Fig 1-9: General outline of the probable propagation of XNA biosynthesis inside the bacterial cells. ALG = alternative leaving group.

The first *in vivo* invasion of XNA replicons is more likely to take place in bacteria, in view of their relatively simple genetic architecture. However, XNAs synthesis is thought to entail high intracellular consumption (*in vivo*) of unnatural building blocks, which will thus need to be provided to the host cell, as XNA nucleotides do not occur in nature. As those molecules are highly polar and remains charged at physiological pH, their passive diffusion in bacterial cells is restrained by hydrophobic lipid bi-layer. On the other hand, most modified nucleosides can enter the bacterial cell easily but are poorly recognized and phosphorylated by host kinases. The bacterial cells will be equipped in turn with suitable enzymatic machinery for polymerization of the monomers. Thus far, efforts have focused on overexpression of endogenous kinases e.g. from *Drosophila Melanogaster*, to broad enzyme-substrate specificity to activate the modified nucleosides to their mono-phosphate form following nucleoside salvage pathways,[59] and more recently, on the transfection of *E. coli* with a plasmid encoding an algal nucleotide triphosphate transporter, which efficiently imports both the triphosphates of d5SICS and dNaM.[26] Despite the ground-breaking value of this last approach, which resulted in the first in vivo replication of an unnatural DNA base pair, its universal applicability to all unnatural bases and types of backbone modifications is yet to be validated.

1.5. Solid phase and liquid phase peptide synthesis

Peptides are chemically synthesized by coupling the carboxylic acid group of one amino acid to the amine group of the next amino acid. Contrary to nature, for peptides are prepared beginning with a pre-activated *C*-terminus and terminates with a *N*-terminus, with suitable protections in between to

avoid undesired by-products, by using solid or solution phase methods. Solid phase peptide synthesis (SPPS) is nowadays a routine technique, developed by R.B. Merrifield almost 60 years ago on a polystyrene based solid phase.[60] SPPS provides a great advantage for the synthesis of large peptides giving overall high yields. In the last few decades the introduction of new polymer supports (binding both carboxyl and amine functions), coupling reagents, and improvements in the protection-deprotection chemistry provided a new dimension to SPPS.[61] [62] The illustrative cycle for solid phase peptide synthesis is showed in Figure 1-10. Besides SPPS, liquid phase peptide synthesis (LPPS) is also sometimes helpful for small peptide up to five-mer. Scalability is an issue with SPPS in an inappropriately equipped laboratory, which can be solved using LPPS for short peptides. Sometimes, the harsh condition (e.g. conc. TFA) used for the final cleavage of peptides from the resin and withstanding of various sensitive orthogonal side chain protections for further functionalization, can be a concern, which also can be addressed by LPPS in a stepwise manner.

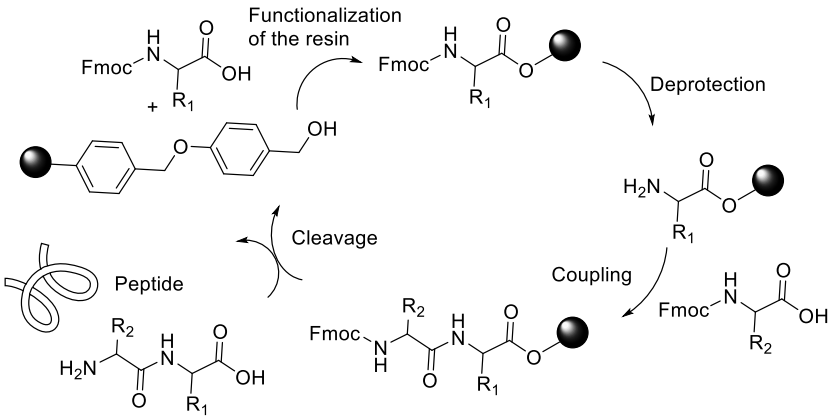


Figure 1-10: Overview of solid phase peptide synthesis using Wang resin.

1.6. Synthesis of 5'-functionalized nucleotides

In the past two decades, different strategies have been developed for the synthesis of nucleotide analogues and phosphoester derivatives as prodrugs for therapeutic purposes. Originally, the synthesis of phosphotriester derivatives was proposed for masking the negative charges of monophosphates to improve lipophilicity and chemical stability. Among the various strategies developed for the synthesis of symmetrical and mixed phosphoesters, the H-phosphonate approach, the phosphite triester method and the use of activated P^V intermediates are the most important (fig. 1-11). The H-phosphonate approach has been successfully used by the Kraszewski group for the synthesis of bis(aryl) nucleoside phosphotriesters.[63] The first step is the activation of an H-phosphonate monoester by diphenyl chlorophosphate or pivaloyl chloride, then coupling with a hydroxyl functionality, followed by in situ oxidation with iodine or NCS, where a second hydroxyl functionality acts as nucleophile to form phosphoesters. The phosphoramidite approach is perhaps the most popular,[64] using a bis(alkyl)-*N,N*-diisopropylphosphoramidite and the corresponding alcohol to

access the pro-moiety. Then, acid catalysed phosphoramidite coupling with a nucleoside in the presence of a promoter like 1-H terazole followed by in situ oxidation (with proper choices among different activating agents or oxidation conditions) yields phosphotriesters. One-pot $POCl_3$ mediated synthesis (P^V chemistry) is one of the popular choices for preparing triphosphates and especially phosphotriesters which are sensitive to the final oxidation step.[65]

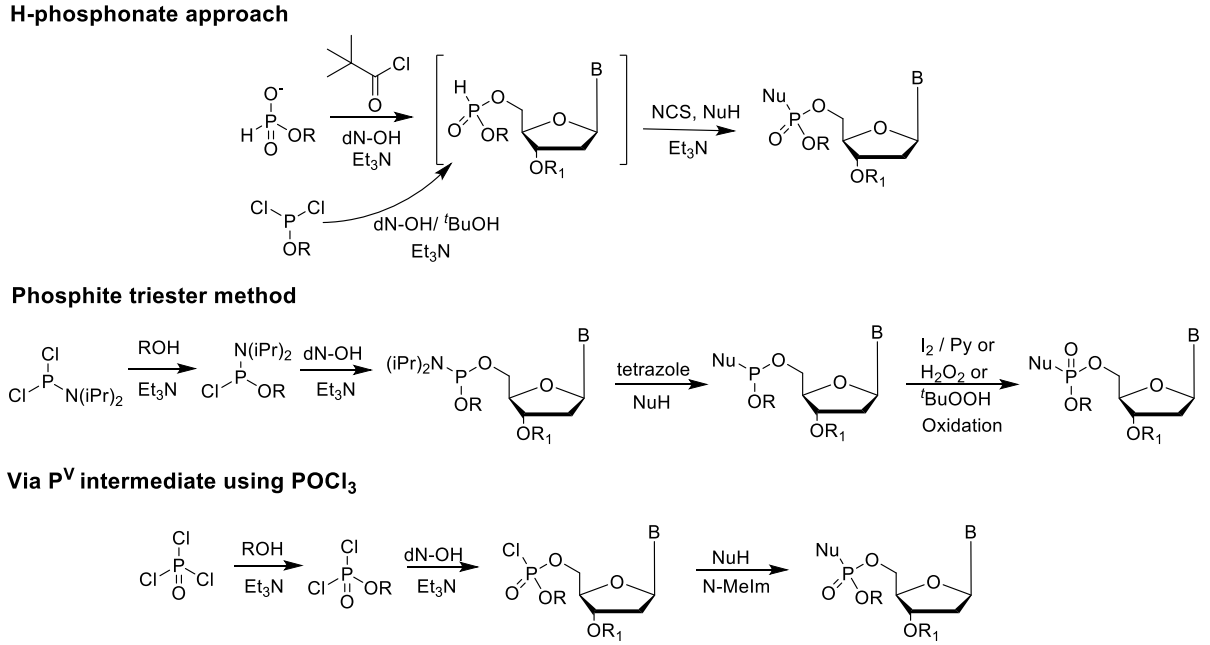


Figure 1-11: Overview of different synthetic routes for the synthesis of phosphate derivatives.

Similarly, phosphoramidates can also be accessed in different ways following diverse strategies that were used for the synthesis of phosphotriesters (fig. 1-12). DCC mediated coupling of a phosphate group of a dNMP with the corresponding amine is still one of the more convenient ways for phosphoramidate synthesis, as reported by Khorana for triphosphate synthesis in 1958.[66] DCC mediated coupling was also used by Abraham et al. to synthesize amino acid-containing phosphoramidates.[67] Using of this method allows the synthesis of pre-activated species for nucleophilic coupling, e.g. imidazolide-dNMP or morpholine-dNMP, that are stable and storable for convenient usage. An alternative to the synthesis of these intermediates is the Mukayiama oxidation-reduction condensation by using dipyridinium disulfide and triphenylphosphine.[68] H-phosphonates, phosphoramidite approach or $POCl_3$ chemistry (described earlier for phosphotriester synthesis) are the other alternatives for the phosphoramidates synthesis using corresponding amine instead of alcohol.

Phosphorothiate and thiophosphoramidate (fig. 1-13) derivatives can be synthesized by using thiophosphoryl chloride or commercially available *O*/*N*-alkylthionophosphodichloridates and coupling the corresponding alcohol / amine in a sequential addition in the presence of a base at low

temperature.[69] Alternatively, such analogues have been also synthesized via the 1,3,2-oxathiaphospholane chemistry introduced by Stec.[70]

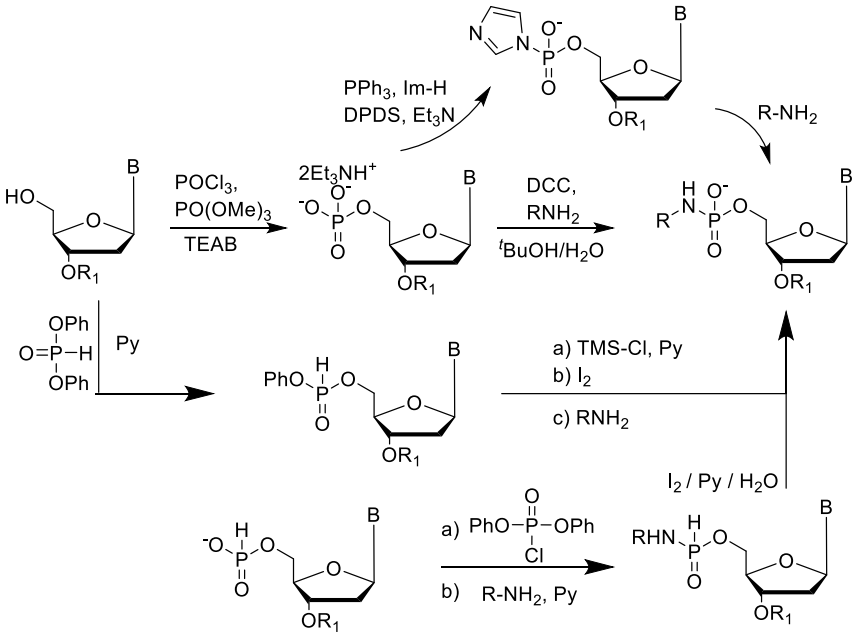


Figure 1-12: Overview of different synthetic routes for the synthesis of phosphoramidate derivatives.

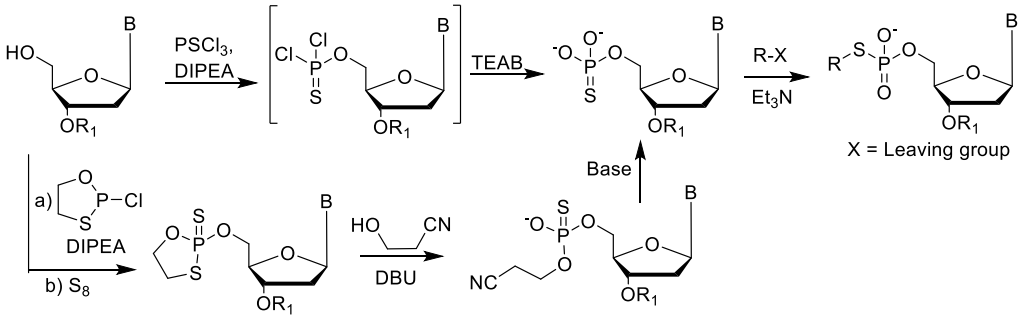


Figure 1-13: Synthesis of phosphorothioate derivatives using thiophosphoryl chloride or via 1,3,2-oxathiaphospholane.

1.7. Conjugation chemistry

Selection of a suitable linkage is an important measure for the successful delivery of any conjugate *in cellulo*. Firstly, the linker should be stable in the extracellular medium and secondly, once the conjugate enters the targeted cells, it should release the attached cargo, after either chemical or enzymatic cleavage. Different types of bio-cleavable linkers have been designed e.g. thioether, disulfide, oxime, carbamate, ester, amide, oxyamide, thiazolidine, hydrazone, urea, phosphate and triazole etc. to enhance cellular uptake and target oligonucleotide delivery.[71] Various chemical strategies for the synthesis of linkers for conjugation to a cargo are described in fig. 1-14. The thioether linkage is one of the most commonly used linkages for conjugation because of its higher

stability. Formation of such linkage occurs either by nucleophilic substitution[72] or by Michael addition[73] of thiols. A disulfide linkage was also utilized for conjugation of nucleotides at the 5'- or 3'- positions, either by direct oxidation of the two reacting thiols[74] or by activating one of the thiol group with 2-2'-dipyridyldithiol followed by nucleophilic substitution by the other thiol group.[75] Along with the numerous reports on the use of amide linkages, research efforts have been also put to develop covalent conjugates with oxime, hydrazone, thiazolidine, by ligation of amino-oxy, hydrazine and cysteine containing peptide respectively with a suitable aldehyde derivative.[76] By utilizing 'click chemistry' between an alkyne and an azide in presence of a Cu (I) catalyst for the formation of 1,2,3-triazoles can also been included in the list of 'linkage' modification.[77] The formation of urea, carbamate or carbonate linkages was also realized by activating amino or hydroxyl functionalities with an activator like carbonyl diimidazole, and then reacting with another fragment bearing an amine or a hydroxyl group.[78]

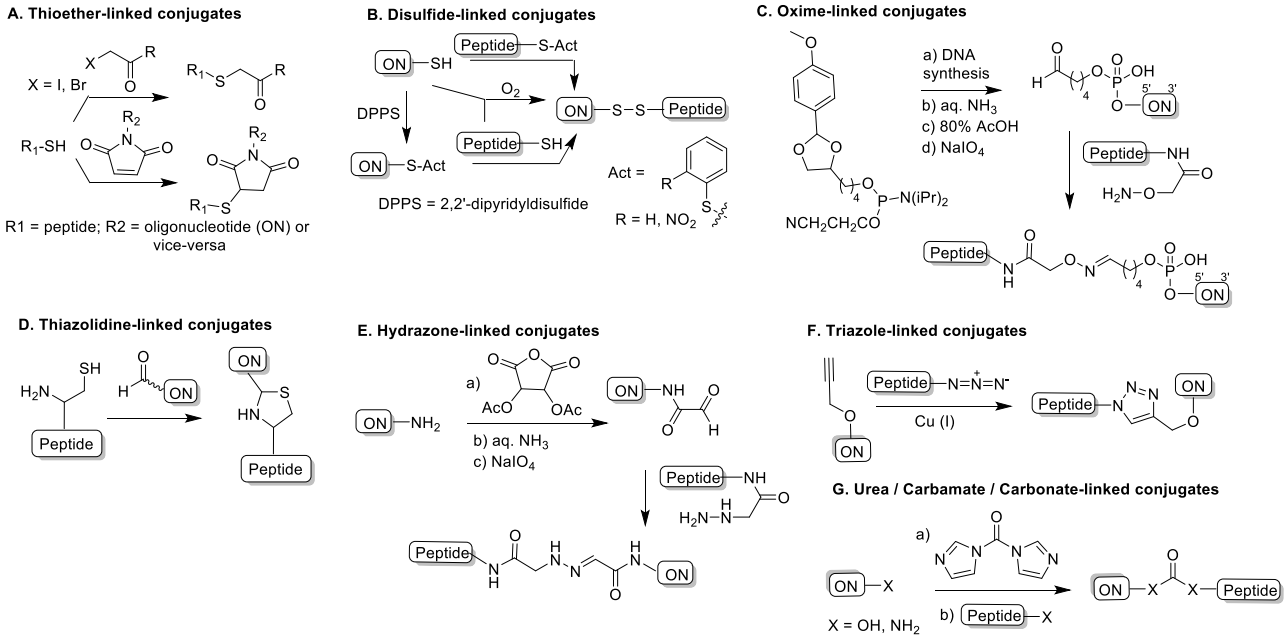


Figure 1-14: Overview of the synthesis of different conjugates with various linkages.

1.8. Microbial transport systems-a potential tool for nucleotides delivery

Both Gram-negative and Gram-positive bacteria possess a rigid cell structure, where the cell membrane is the most dynamic element in the cell. The discovery of novel antibacterial drugs for combating resistance issues is a very hard challenge due to the limited permeability of many drugs or to the efflux systems of bacteria which can expel the drug outside the cell. The cell wall of both Gram-negative and Gram-positive bacteria is composed of a multilayered rigid structure to shield the cell from osmotic shock and physical damage. In most Gram-positive bacteria, the cell wall consists

of thick multilayered peptidoglycans and thin lipids linked with teichoic acid. On the contrary, the cell walls of Gram negative bacteria are thinner, composed of a thin peptidoglycan layer and an outer membrane comprising lipopolysaccharides (LPS), lipoproteins and phospholipids. This is the key barrier for transport of charged polar molecules and many antibiotics. Gram negative bacteria also possess a periplasmic space which contains many enzymes for transport, degradation and synthesis. Generally, water and few small uncharged molecules (Mol. Wt.< 100 Daltons) can enter the bacterial cell through membrane porins, but the bacterial membrane does not allow the passage of large substances and charged molecules, whilst charged nutrients (e.g. Na^+, K^+) are transported by transport proteins present in the membrane via specific transport systems.

Mostly the membrane proteins which are involved in transport processes are designated broadly as porters, transport systems, carrier proteins and permeases. Mainly three types of transport processes are involved which are known as uniport, symport and antiport. Uniporters allow a solute to pass through in one direction, symporters allow the passage of two solutes in the same direction and time, and finally antiporters allow two solutes to pass in opposite directions at the same time. Bacteria have evolved a compact transport system utilizing various transport proteins, due to their urge to concentrate essential nutrients inside the cytoplasm against the concentration gradient of the environment, in different extreme conditions. There are mainly four well-accepted categories of transporters, a) channels, b) porters, c) primary active transporters and d) group translocators. Generally the facilitated diffusion system and passive transport of molecules or ions across the membranes are aided by channels, which are less common in bacteria. Accumulation of solutes inside bacterial cells needs active transport systems, which can be of two types, ion driven transport systems (IDT) and binding-protein dependent transport systems (BPDT). Through IDT, many amino acids and ions are accumulated via symport or antiport processes utilizing energy from proton motive force (pmf) or chemiosmotic potential. As the bacterial active transport machinery accumulates solutes inside the cytoplasm at higher concentrations than that of the environment, which requires energy according to law of chemical kinetics. Generally the primary active transporters use a 'primary' source of energy either from chemical, electrical or solar sources. There are various types of active transporter systems differing by the way of utilizing sources of energy e.g. P-P bond hydrolysis-driven transporters (ATP hydrolysis), decarboxylation driven transporters, methyl transfer-driven transporters, oxidoreduction-driven transporters and light absorption-driven transporters. On the other hand, binding-protein dependent transport systems (BPDT) are far more complex than the facilitated transport, as they comprise four proteins. Out of these four proteins, two forms a channel in the transmembrane domain that allows passage of the solute, the third protein is located in the periplasmic space and has a very strong binding affinity for the specific solute, translocating the solute to a fourth protein which acknowledges the solute into the membrane channel. A conformational change allows the solutes through the channel by using of energy generally provided by the hydrolysis of ATP to

ADP. One of such subfamily of active transporter is the ABC (ATP Binding Cassette) transporter, where the fourth protein is the nucleotide binding domain (NBD) and ATP hydrolysis is linked to the active transport.

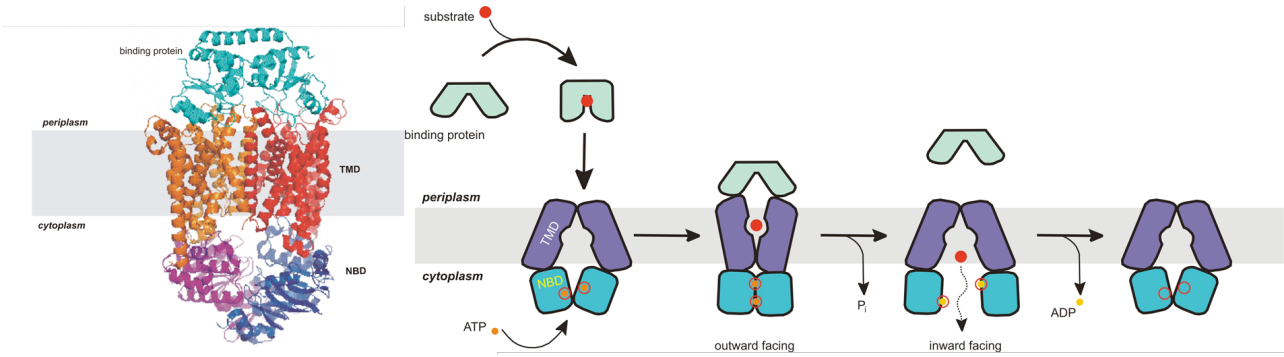


Figure 1-15: Proposed mechanism of transport of ABC transporters (reproduced from https://www.boundless.com/microbiology/textbooks/boundless-microbiology-textbook/bacteria-archaea-and-eukaryote-cell-structure-4/transport-across-the-cell-membrane-33/abc-transporters-256-1009/)

In bacteria, ABC transporters[79] transport a broad range of substrates including peptides, sugars, ions, amino acids, vitamins like biotin and thiamine, sulfonates and sulfates and other molecules that are mostly hydrophilic, which otherwise would have a restricted entry from the lipid bi-layer of the membrane. However for the nucleotides there is no such specific transporter in bacteria, and as nucleotides are also charged hydrophilic compounds at physiological pH their entry is also restricted by lipid bi-layers. The exploitation of one of these transporters by constructing nucleotide conjugates with suitable solutes could make it possible to deliver these molecules in the bacterial cell. The systematic exploration of these transporter systems, may pave the way for nucleotide delivery. Due to the wide range of possibilities to assemble nucleotide conjugates, only some of them were explored in the limit of this thesis, which are described in the following paragraphs.

1.8.1. Peptide transporters in microorganisms

Extensive research has been performed by various groups on the structure, physiological role and molecular mechanism of different peptide permeases from *E. coli* and *S. typhimurium*.[80] Three types of peptide permeases are mainly found in bacteria: the dipeptide (Dpp), tripeptide (Tpp) and oligopeptide permease (Opp). Recent research[81] based on molecular recognition, unveiled that they are complementary in many ways; a) for a particular peptide each recognizes different conformational forms, b) capable to transport all small peptides from the peptide pool by operating simultaneously. Specificity and selectivity of substrate transport via Dpp has been determined by several techniques like radio- or fluorescent-labelled peptides utilizing Opp and Tpp auxotrophic *E.coli* strain[80], which revealed that although Dpp can transport both di- and tri-peptides, the rate is much higher for di-

peptides. Like Dpp, Tpp can also transport both di- and tri-peptides, but with a different substrate specificity. Amongst the three peptide permeases, generally Opp is the most abundant periplasmic protein (7-10% of total periplasmic proteins) and structurally well-studied.[82] It was crystallized in its free form and also with ligated peptides, its three dimensional structure by X-ray analysis revealed three-domains and the existence of large hydrated cavities for accommodating the lateral chains of amino acids in order to cope with the various peptide structures that the transport system can be confronted with during the natural life of *E. coli*. Remarkably, upon binding to diverse ligands and chemical groups, the structure of the protein is not affected.

Table 1-1: Optimal features for molecular recognition of peptides by peptide permeases Dpp, Tpp and Opp.

Feature	Dpp-Type	Tpp-Type	Opp-Type
Number of residues	2 (3)	2 (3)	(2) 3-6
N-terminal α-amino group	Positively charged amino group optimal; substitution with charge retention acceptable		
C-terminal α-carboxyl group	Negatively charged carboxylate optimal		charged carboxylate not essential
Peptide bonds	All trans (± 180º) highly preferred		
Residue chirality	All L-stereo chemistry optional		
N-C distance (distance betwwen N- and C-terminus)	5.1-6.4 Å	4.5-5.6 Å	> 6.5 Å
Side chains	All natural and some larger, unnatural side chains recognized		
Isopotential field	There should be symmetric about the N- and C- termini		
Stabilization by backbone atoms	Unmodified peptide bond NH and CO groups are ideal to H-bond with suitable groups on transporter proteins		

Different features of the each permease are summarized in Table 1-1, which have been established using a range of biological and biophysical assays. *In vitro* binding studies on peptide complexes of *E.coli* OppA revealed high affinity and broad substrate specificity for small peptides comprising two up to five/six L-amino acids, regardless of their sequence. It was also revealed that OPP has a preference for peptides containing Ala, Gly, Phe[83] and basic residues[84], specifically

lysine, which may result from a negatively charged surface in proximity of the initial binding site. From different investigations it has been established that the presence of free α-amino and carboxyl groups at the *N*- and C-terminal respectively was important for binding to *E.coli* OppA, but it remained impossible to explain why other protected peptides were also translocated, though less effectively.[83] By exploiting peptide-delivery systems different therapeutic[85] and antimicrobial [86]compounds were also transported inside the bacterial cell, which can be challenging to be delivered by direct means. To the best of our knowledge, peptide delivery system has never been exploited for nucleotide delivery in bacterial cell. Although there are reports on the use of cell penetrating peptides (CPP) as a vehicle for cargo delivery in mostly mammalian cells, these CPPs also exhibit antimicrobial properties limiting their use in bacterial cells for cargo delivery.[87]

1.8.2. Biotin transporter in microorganism

Biotin is an important nutrient for many bacteria and most of them are capable of *de novo* biosynthesis of biotin. Several of them are also able to transport exogenous biotin for their growth when biotin is present in the medium[88] and even concentrations of 5 ng/mL are sufficient for bacterial growth as first reported by Pai and Lichstein using a biotin auxotrophic *E. coli* strain.[89] After this, several groups have tried to identify gene coding for biotin transporters both for gram negative and gram positive bacteria.[90] It is well accepted that energy-coupling factor (ECF) transporters, a special type of ABC transporters, are responsible for biotin transport.[91] In most bacteria the BioMNY module is found, but in *E. coli* an analogous biotin transporter YigM was identified. Moreover, biotin transporter systems were also exploited for cargo delivery in bacterial cells. Walter and Altmen reported that biotinylated peptide conjugates can be delivered up to 31-mer peptide sequence in *E. coli* cells, which is otherwise limited to a maximum of five to six-mer for non-biotinylated peptides.[92] Although not reported for uptake studies, a variety of chemically conjugated biotinylated nucleotides are known for labeling, incorporation, PCR and quantification studies.[93]

1.8.3. Sulfonate transporters in microorganisms

Sulfur is an essential nutrient in all microorganisms for the biosynthesis of crucial building blocks, such as sulfur containing amino acids cysteine and methionine, and cofactors like biotin, thiamine, lipoic acid, molybdoptrein, glutathione, coenzyme A and M. The source of sulphur in bacterial cells is either in inorganic form like sulfates and sulfides or an external organically bound sulfur like sulfonates ($R\text{-}SO_3^-$) and sulfate esters ($ROSO_3^-$). Under sulfur starvation conditions bacteria express different proteins for transport of external sulfonates and sulfates, which are metabolized through a series of enzymatic steps, giving sulfide as degradative intermediate[94] and eventually sulfur as nutrient, through the sulphur assimilation cycle.[95] As sulfates, sulfonates and sulfonate esters are charged polar molecules with low pK_a values at physiological pH, their passive

diffusion is hindered by the lipid bi-layer, and rather involves an active uptake system for cellular accumulation.[96] In bacteria, the role of active transport for sulfonates and sulfates accumulation has been extensively studied. It was revealed that the *tau*ABC protein, encoded by the *tau*ABCD gene cluster and *ssu*ABC, encoded by the *ssu*EADCB gene cluster, are responsible for taurine and alkanesulfonates uptake respectively.

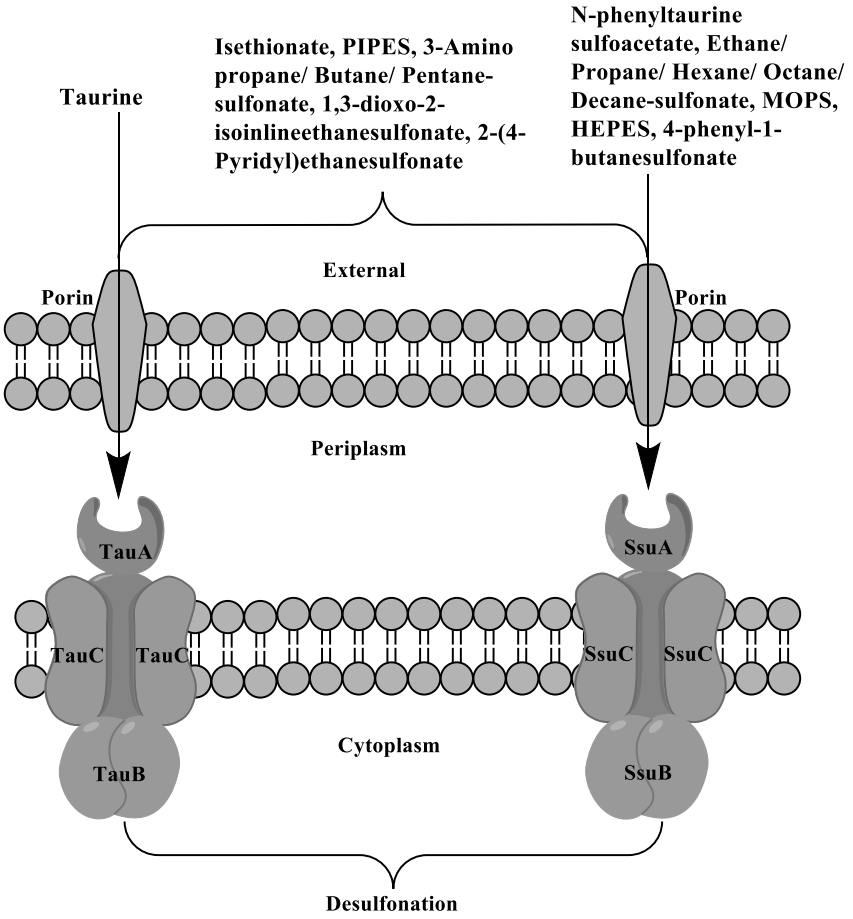


Figure 1-16: Substrate specificities of *E.coli Tau*ABC and *Ssu*ABC transporters (partly adapted from Ref:[95])

The TauD gene expresses for the protein alpha-ketoglutarate-dependent dioxygenase, whilst the SsuD gene expresses a flavin mononucleotide dependent monooxygenase, both promoting the desulfonation process from taurine and a wide range of other sulfonates respectively.[95] TauABC and ssuABC also belong to the ABC type transport systems and like other ABC transporters they also consist of specific periplasmic binding proteins and consume energy from ATP hydrolysis. A series of sulfonate compounds were evaluated as a source of sulphur, as summarized in fig. 1-16, and the specificity of the transporters was also determined by growing phenotypes of different *E.coli* mutants devoid of one or more genes for the putative TauABC and SsuABC transporters. It was observed that few particular sulfonates entered the cell exclusively via a specific transporter whilst few others were

taken up by both the transporters.[95] Although there are no reports of indicating that sulfonate transporters could be used for cargo delivery inside bacterial cells, their potential use for cellular delivery within a sulfur-based nutritional selection system cannot be excluded.

1.8.4. Aminoglycosides uptake in bacteria

Aminoglycosides (AGs) are clinically important drugs, showing broad-spectrum antibacterial activity against both gram-negative and gram-positive bacteria, binding to the bacterial ribosomal 30S and 50S subunits and inhibiting protein synthesis. The uptake of aminoglycosides inside the bacterial cells is an energy requiring process and can be accumulated against a concentration gradient. The detailed mechanism by which AGs penetrate into the bacterial cytoplasm remains obscure but based on available models (streptomycin and gentamicin) it has been postulated that cellular uptake of AGs comprised of three different stages.[97] The first step is the non-specific electrostatic interaction between positively charged aminoglycosides and negatively charged lipopolysaccharide (LPS) or phospholipids of the bacterial outer membrane followed by two energy dependent stages named energy-dependent phase I (EDPI) and energy-dependent phase II (EDPII). The first step is reversible, concentration dependent and unaffected by the inhibitors of the subsequent energized steps. In case of EDPI, the rate of uptake is dependent on the external concentration of AGs and also prone to inhibitors of oxidative phosphorylation and electron transporter. Although the exact mechanism in EDPII remains elusive, it is thought to include an accelerated energy-dependent transport of AGs across the cytoplasmic membrane following EDPI, through a process that uses energy from electron transport and ATP hydrolysis and which can be inhibited by some inhibitors of protein synthesis. For their cationic nature and strong binding properties with negatively-charged species, AGs were used in micelle formation and as cationic lipids for gene transfection, amongst others.[98] Although several modes of bacterial resistance to AGs have been described, enzymatic modification is the most common cause of resistance,[99] which occurs intra-cellularly. In most cases AGs penetrate bacterial cell membrane effectively, leaving the option of utilizing AGs in bacterial cargo delivery, especially in resistant mutants. Due to the lipophilic nature of polyamines, several AGs[100] and guanidinoglycosides[101] based conjugates have been introduced to enhance cellular uptake and transfection for poly-nucleic acid drugs (siRNA or gene delivery), mainly in mammalian cells.

1.9. Thymineless death (TLD)

Nucleotides are constructed from small natural fragments that go through a sequence of biochemical transformations promoted by specific enzymes mainly following de novo pathways in bacterial cells (Fig. 6-1). In de novo biosynthesis of dTTP, the enzyme thymidylate synthase plays a crucial role, catalysing the methyl group transfer from 5,10-methylenetetrahydrofolate to the 5-position of the uridine ring in dUMP leading to the formation of dTMP. The salvage pathway for

deoxyribinucleotides involves conversion of nucleosides to their monophosphate derivatives that are recovered during the degradation of DNA and RNA or supplied externally. In *E. coli*, thymidine kinase converts thymidine to its mono-phosphate form (Fig. 1-17).

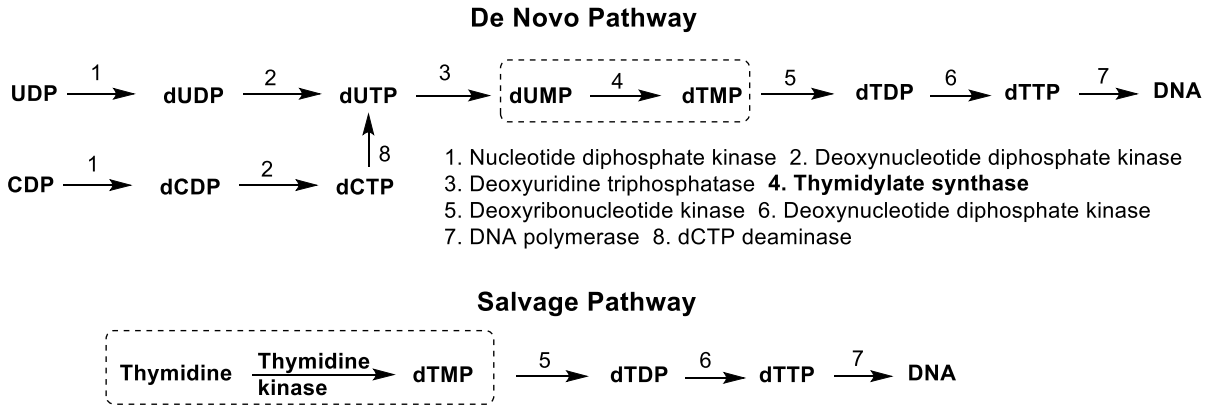


Figure 1-17: Thymidine triphosphate metabolism in *E. coli*.

Under thymine starvation conditions microorganisms suffer from cell death a phenomenon known for many years as thymineless death (TLD), on the contrary other nutritional requirements having a biostatic rather than deadly consequences. Thymine starvation has both direct and indirect effects, which encompasses both single- and double-stranded DNA breaks respectively. Although recombinational DNA repair processes could restore the single-stranded DNA damage, the double-stranded DNA breaks ultimately leading to cell death. The deletion of the *thyA* gene in *E. coli* strains which encodes for thymidylate synthase causes thymine starvation and ultimately leads to TLD, as first observed by Cohen and Barner using a thymine auxotrophic mutant (*thyA*-).[102] No active thymidylate synthetase is functional in thymine auxotrophic *E.coli* (*thyA*-) cells; thus as a result no dTMP, dTDP or dTTP are produced endogenously. Unlike wild type cells, *thyA* mutants can utilize exogenous thymine for the production of dTTP via a salvage pathway. The salvage pathway involves deoxyribose-1-phosphate (dRib-1-P) and thymine, which are converted first to thymidine and subsequently to dTMP by the action of thymidine kinase. The lack of dRib-1-P pool in wild type strain generally explains the reason of their inability to use exogenous thymine. Because of the absence of thymidylate synthetase, the conversion of dUMP to dTMP is prevented. The cumulated dUMP then degrades to uracil and dRib-1-P thus acting as the source of dRib-1-P pool in *thyA* mutants. Apart from gene deletion, research efforts have been focused on developing other approaches to achieve thymine starvation e.g. (a) Inhibiting folate metabolism by using either sulphonamides,[103] cytosine arabinoside,[104] showdomycin[105] or methotrexate / aminopterin / trimethoprim (dihydrofolate reductase inhibitor),[106] and thereby interfering with thymidylate synthesis (b) Using FUdR a direct inhibitor of thymidylate synthase (c) Inhibiting thymidine

phosphorylase that converts thymine to thymidine with addition of a ribonucleoside in the growth medium.[107]

1.10. Objectives of the research

In view of nucleotides low cell-penetrating ability and inclination to undergo fast enzymatic dephosphorylation before being taken up, we wished to develop a delivery tool for their active uptake, based on a nutritional selection system. Each organism needs to take up extracellular nutrients in order to ensure the necessary supply of materials and energy for cellular functioning. Various transport systems are known to carry out the transfer of organic and inorganic molecules from outside the cell into the cytoplasm.

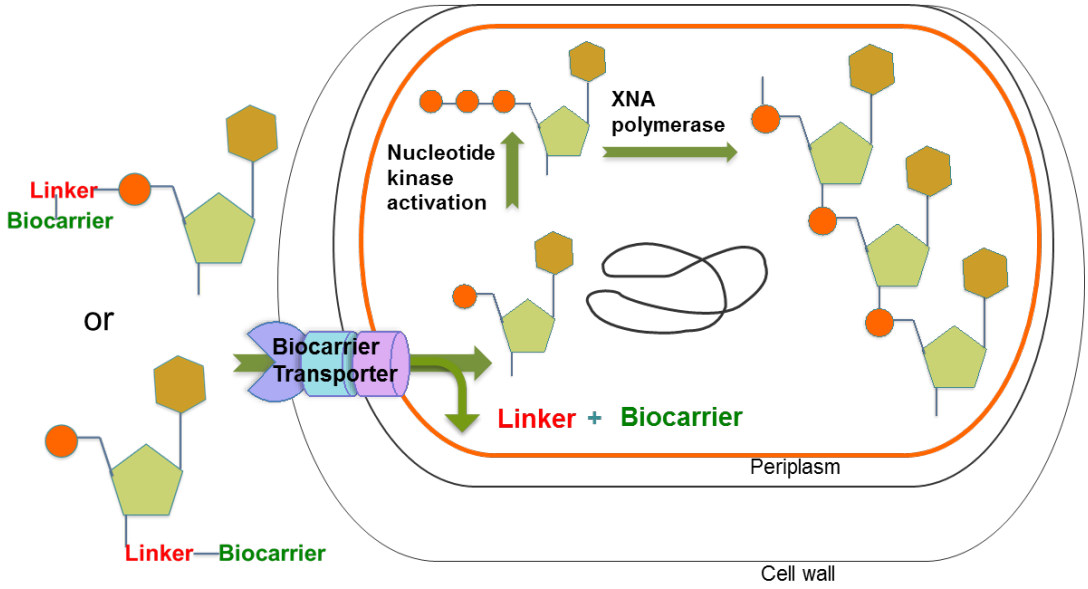


Figure 1-17. Concept for the rational design of peptide-nucleotide conjugates as delivery systems of natural and modified nucleotides in prokaryotic cells.

The general concept of our proposed approach consists in linking the information system to a nutrient essential for the cell, thus taking advantage of natural transport pathways or uptake systems occurring in bacterial cells, to eventually facilitate nucleotide delivery across the cytoplasmic membrane. A bio-cleavable covalent bond is designed to maintain the conjugate intact during cellular uptake, but allow facile release of the nucleotide intracellularly (either chemically or enzymatically). Building on previous studies, 2'-deoxythymidine monophosphate (dTMP) is selected as archetypal nucleotide system. The relevant dTMP conjugate is supplied within the nutrient medium to a thymidine auxotrophic *E.coli* strain deleted for the ThyA gene encoding for thymidylate synthase, which is known to unconditionally require thymine or thymidine for growth. So in principle the engineered auxotrophic strains can only survive only if dTMP is transported and released inside the

bacterial cells. Based on this model study and upon successful delivery of dTMP inside the bacterial cells, the vectors for the active uptake system can be fabricated for XNA delivery mimicking similar transport mechanisms.

It is the aim of this thesis to create a variety of biocarrier-nucleotide conjugates for uptake evaluation in auxotrophic bacterial cells. We plan to perform the initial work in *E. coli*, since, it represents an ideal model organism for studying cellular processes, as its biology is well known. With this background, we set our objectives to synthesize and evaluate a diverse set of biocarrier-nucleotide conjugates, linked with a labile linker, and conjugated either at the 3'-hydroxyl or at the 5'- phosphate position of dTMP. Upon successful foundation of the auxotrophic model, particular delivery vehicle can also be attached to nucleotides containing fluorescent bases, which can be visualized under confocal microscopy as an additional evidence of this hypothesis.

Di- to hexapeptides containing L-amino acids are easily taken up by bacterial cells irrespective of their sequences. The structural basis for molecular recognition of such substrates by peptide transporters has been extensively studied (see section 1.8.1). Few examples of the successful exploitation of peptide transporter systems for delivery of different antimicrobial and therapeutic agents are also reported, but this concept has never been expanded to bacterial uptake of nucleotide analogues. A series of oligopeptides coupled via a bio-labile linker to either in the 3'- or 5'-position of TMP has been developed aiming at nucleotides delivery in bacterial cells. The efficient synthesis and purification of different macromolecules of peptide-nucleotide conjugates (PNCs), and biocarrier-conjugates in general are a very challenging task in view of the introduction of highly polar phosphate groups concurrently with a bio-labile linker possessing a modest stability. The development of new robust and improved synthetic methods has been planned to obtain different biocarrier-nucleotide conjugates in high purity and yield. In the first project (**Chapter 2**) we aimed at the synthesis and evaluation of new PNCs linked at the 3'-hydroxyl group of dTMP for nucleotides delivery in bacterial cells exploiting naturally occurring peptide permeases. Five different linkers were chosen for conjugation of PNCs which are assumed to be liable to cleavage inside the bacterial cell i.e. carbamate (OCONH), ester (OCO), oxyamide (ONHCO), oxymethyleneoxyamide (OCH_2ONHCO), oxymethylene-oxyester (OCH_2OCO). Different α-substituted glycine containing di- and tri-peptides were described for delivering and releasing the attached cargo inside the bacterial cells, as first reported by Gilvarg *et al*. In the second project (**Chapter 3**), synthesis and evaluation of diverse novel 5'-phospho-linked peptide nucleotide conjugates (PNCs) assembled from α-substituted glycine dipeptides, were planned for nucleotides delivery studies in bacterial cells. Different synthetic routes and metabolic stability studies were designed for these PNCs with the two assembling units, linked by various phosphoesters linkers such as phosphate, phosphoramidate, phosphorothioate and

thioethyl-phosphate between the α-carbon of a C-terminal glycine residue and the 5'-hydroxyl group of thymidine. Many bacteria are also capable of transporting extracellular sulfonates ($R\text{-}SO_3^-$) and sulfate esters ($ROSO_3^-$) inside the cells as a source of sulfur under sulphur-starvation conditions. In the third project (**Chapter 4**) a series of sulfonate derived phosphoramidate conjugates of dTMP, connected through 5'-hydroxyl group of dTMP, were designed and synthesized, and studied both as active intermediates in the enzymatic primer-extension of DNA and for nucleotide delivery in bacteria targeting sulfonate transporters (see section 1.8.3). Generally bacteria are able to transport peptides up to six-mers, but biotinylation of peptides facilitates transportation of peptide sequences up to 31-mer in bacterial cells targeting biotin transport systems (see section 1.8.2). Due to the lipophilic nature of polyamines, aminoglycosides are well known for their bacterial cell-penetrating ability (see section 1.8.4). In the fourth project (**Chapter 5**) different biotin-nucleotide and aminoglycoside-nucleotide systems (linked at 3'-hydroxyl of dTMP) were designed, synthesized and evaluated for cellular delivery in *E.coli* auxotrophic strains, targetting the corresponding biotin transporters or aminoglycoside uptake systems.

1.11. References

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737-738.
[2] M. Sundaralingam, *Journal of the American Chemical Society* **1965**, *87*, 599-606.
[3] B. Basham, G. P. Schroth, P. S. Ho, *Proceedings of the National Academy of Sciences* **1995**, *92*, 6464-6468.
[4] W. Szybalski, *Adv Exp Med Biol* **1974**, *44*, 23-24.
[5] M. Rezaei, S. H. Zarkesh-Esfahani, *Journal of Research in Medical Sciences* **2012**, *17*, 681-685.
[6] L. Villakomaroff, A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S. P. Naber, W. L. Chick, W. Gilbert, *Proceedings of the National Academy of Sciences of the United States of America* **1978**, *75*, 3727-3731.
[7] Y. Kung, W. Runguphan, J. D. Keasling, *Acs Synthetic Biology* **2012**, *1*, 498-513.
[8] A. Prindle, P. Samayoa, I. Razinkov, T. Danino, L. S. Tsimring, J. Hasty, *Nature* **2012**, *481*, 39-44.
[9] D. L. Porter, B. L. Levine, M. Kalos, A. Bagg, C. H. June, *New England Journal of Medicine* **2011**, *365*, 725-733.
[10] P. Q. Nguyen, Z. Botyanszki, P. K. R. Tay, N. S. Joshi, *Nature Communications* **2014**, *5*.
[11] H. Bugl, J. P. Danner, R. J. Molinari, J. T. Mulligan, H.-O. Park, B. Reichert, D. A. Roth, R. Wagner, B. Budowle, R. M. Scripp, J. A. L. Smith, S. J. Steele, G. Church, D. Endy, *Nature Biotechnology* **2007**, *25*, 627-629.
[12] P. Ronald, *Genetics* **2011**, *188*, 11-20.
[13] A. Eschenmoser, *Science* **1999**, *284*, 2118-2124.
[14] S. A. Benner, *Accounts of Chemical Research* **2004**, *37*, 784-797.
[15] A. M. Sismour, S. A. Benner, *Nucleic Acids Research* **2005**, *33*, 5640-5646.
[16] Z. Yang, D. Hutter, P. Sheng, A. M. Sismour, S. A. Benner, *Nucleic Acids Research* **2006**, *34*, 6095-6101.
[17] Z. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, S. A. Benner, *Nucleic Acids Research* **2007**, *35*, 4238-4249.
[18] A. T. Krueger, E. T. Kool, in *Chem Biol, Vol. 16*, United States, **2009**, pp. 242-248.
[19] M. Winnacker, E. T. Kool, *Angewandte Chemie International Edition* **2013**, *52*, 12498-12508.

[20] S. R. Lynch, H. Liu, J. Gao, E. T. Kool, *Journal of the American Chemical Society* **2006**, *128*, 14704-14711.
[21] J. C. Delaney, J. Gao, H. Liu, N. Shrivastav, J. M. Essigmann, E. T. Kool, *Angewandte Chemie International Edition* **2009**, *48*, 4524-4527.
[22] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat Meth* **2006**, *3*, 729-735; I. Hirao, T. Mitsui, M. Kimoto, S. Yokoyama, *Journal of the American Chemical Society* **2007**, *129*, 15549-15555.
[23] M. Kimoto, R. Kawai, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Research* **2009**, *37*.
[24] K. Dhami, D. A. Malyshev, P. Ordoukhanian, T. Kubelka, M. Hocek, F. E. Romesberg, *Nucleic Acids Research* **2014**, *42*, 10235-10244.
[25] L. Li, M. Degardin, T. Lavergne, D. A. Malyshev, K. Dhami, P. Ordoukhanian, F. E. Romesberg, *Journal of the American Chemical Society* **2014**, *136*, 826-829.
[26] D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Correa, F. E. Romesberg, *Nature* **2014**, *509*, 385-388.
[27] M. Petersen, J. Wengel, *Trends in Biotechnology* **2003**, *21*, 74-81.
[28] K. E. Nielsen, J. Rasmussen, R. Kumar, J. Wengel, J. P. Jacobsen, M. Petersen, *Bioconjugate Chemistry* **2004**, *15*, 449-457.
[29] R. N. Veedu, J. Wengel, *Molecular Biosystems* **2009**, *5*, 787-792; R. N. Veedu, B. Vester, J. Wengel, *Journal of the American Chemical Society* **2008**, *130*, 8124-+.
[30] K. U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347-1351.
[31] M. O. Ebert, C. Mang, R. Krishnamurthy, A. Eschenmoser, B. Jaun, *Journal of the American Chemical Society* **2008**, *130*, 15105-15115.
[32] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S. Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, P. Holliger, *Science* **2012**, *336*, 341-344.
[33] E. Lescrinier, R. Esnouf, J. Schraml, R. Busson, H. A. Heus, C. W. Hilbers, P. Herdewijn, *Chemistry & Biology* **2000**, *7*, 719-731.
[34] C. Hendrix, H. Rosemeyer, I. Verheggen, A. Van Aerschot, F. Seela, P. Herdewijn, *Chemistry – A European Journal* **1997**, *3*, 110-120.
[35] J. Wang, B. Verbeure, I. Luyten, E. Lescrinier, M. Froeyen, C. Hendrix, H. Rosemeyer, F. Seela, A. Van Aerschot, P. Herdewijn, *Journal of the American Chemical Society* **2000**, *122*, 8595-8602.
[36] V. Kempeneers, M. Renders, M. Froeyen, P. Herdewijn, *Nucleic Acids Research* **2005**, *33*, 3828-3836.
[37] G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 4398-4402.
[38] S. Zhang, C. Switzer, J. C. Chaput, *Chemistry & Biodiversity* **2010**, *7*, 245-258.
[39] B. D. Heuberger, C. Switzer, *Journal of the American Chemical Society* **2008**, *130*, 412-413.
[40] M. K. Schlegel, A. E. Peritz, K. Kittigowittana, L. L. Zhang, E. Meggers, *Chembiochem* **2007**, *8*, 927-932.
[41] J. J. Chen, C. H. Tsai, X. Cai, A. T. Horhota, L. W. McLaughlin, J. W. Szostak, *Plos One* **2009**, *4*.
[42] C. H. Tsai, J. Y. Chen, J. W. Szostak, *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 14598-14603.
[43] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497-1500.
[44] R. Gambari, *Expert Opin Ther Pat* **2014**, *24*, 267-294.
[45] C. Bohler, P. E. Nielsen, L. E. Orgel, *Nature* **1995**, *376*, 578-581.
[46] P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm, B. Norden, *Nature* **1994**, *368*, 561-563; K. E. Nelson, M. Levy, S. L. Miller, *Proceedings of the National Academy of Sciences* **2000**, *97*, 3868-3871.
[47] Y. Ura, J. M. Beierle, L. J. Leman, L. E. Orgel, M. R. Ghadiri, *Science* **2009**, *325*, 73-77.
[48] P. Herdewijn, P. Marlière, *FEBS Letters* **2012**, *586*, 2049-2056.
[49] L. A. Alexandrova, A. Y. Skoblov, M. V. Jasko, L. S. Victorova, A. A. Krayevsky, *Nucleic Acids Research* **1998**, *26*, 778-786; R. G. Yount, D. Babcock, W. Ballantyne, D. Ojala, *Biochemistry* **1971**, *10*, 2484-2489; R. S. Goody, F. Eckstein, *Journal of the American Chemical Society* **1971**, *93*, 6252-6257.

[50] O. Adelfinskaya, P. Herdewijn, *Angewandte Chemie International Edition* **2007**, *46*, 4356-4358.
[51] O. Adelfinskaya, M. Terrazas, M. Froeyen, P. Marliere, K. Nauwelaerts, P. Herdewijn, *Nucleic Acids Research* **2007**, *35*, 5060-5072.
[52] A. Giraut, N. Dyubankova, X. P. Song, P. Herdewijn, *Chembiochem* **2009**, *10*, 2246-2252.
[53] A. Giraut, P. Herdewijn, *ChemBioChem* **2010**, *11*, 1399-1403.
[54] S. Yang, M. Froeyen, E. Lescrinier, P. Marliere, P. Herdewijn, *Organic & Biomolecular Chemistry* **2011**, *9*, 111-119; X. P. Song, C. Bouillon, E. Lescrinier, P. Herdewijn, *Chembiochem* **2011**, *12*, 1868-1880; A. Giraut, X. P. Song, M. Froeyen, P. Marliere, P. Herdewijn, in *Nucleic Acids Res, Vol. 38*, England, **2010**, pp. 2541-2550.
[55] H. Pedersen, S. Holder, D. P. Sutherlin, U. Schwitter, D. S. King, P. G. Schultz, *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 10523-10528.
[56] F. J. Ghadessy, J. L. Ong, P. Holliger, *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 4552-4557.
[57] P. Herdewijn, P. Marliere, *Chemistry & Biodiversity* **2009**, *6*, 791-808.
[58] A. I. Taylor, V. B. Pinheiro, M. J. Smola, A. S. Morgunov, S. Peak-Chew, C. Cozens, K. M. Weeks, P. Herdewijn, P. Holliger, *Nature* **2015**, *518*, 427-430.
[59] Y. Wu, M. Fa, E. L. Tae, P. G. Schultz, F. E. Romesberg, *Journal of the American Chemical Society* **2002**, *124*, 14626-14630; S. K. Jarchow-Choy, E. Sjuvarsson, H. O. Sintim, S. Eriksson, E. T. Kool, *Journal of the American Chemical Society* **2009**, *131*, 5488-5494.
[60] R. B. Merrifield, *Journal of the American Chemical Society* **1963**, *85*, 2149-2154.
[61] M. Amblard, J.-A. Fehrentz, J. Martinez, G. Subra, *Molecular Biotechnology* **2006**, *33*, 239-254.
[62] M. Amblard, J.-A. Fehrentz, J. Martinez, G. Subra, in *Peptide Synthesis and Applications, Vol. 298* (Ed.: J. Howl), Humana Press, **2005**, pp. 3-24.
[63] J. Romanowska, A. Szymańska-Michalak, J. Boryski, J. Stawiński, A. Kraszewski, R. Loddo, G. Sanna, G. Collu, B. Secci, P. La Colla, *Bioorganic & medicinal chemistry* **2009**, *17*, 3489-3498.
[64] I. Lefebvre, C. Perigaud, A. Pompon, A.-M. Aubertin, J.-L. Girardet, A. Kirn, G. Gosselin, J.-L. Imbach, *Journal of Medicinal Chemistry* **1995**, *38*, 3941-3950; T. P. Prakash, M. Prhavc, A. B. Eldrup, P. D. Cook, S. S. Carroll, D. B. Olsen, M. W. Stahlhut, J. E. Tomassini, M. MacCoss, S. M. Galloway, C. Hilliard, B. Bhat, *Journal of Medicinal Chemistry* **2005**, *48*, 1199-1210; N. Schlienger, C. Périgaud, G. Gosselin, J.-L. Imbach, *The Journal of Organic Chemistry* **1997**, *62*, 7216-7221.
[65] A.-L. Villard, G. Coussot, I. Lefebvre, P. Augustijns, A.-M. Aubertin, G. Gosselin, S. Peyrottes, C. Périgaud, *Bioorganic & Medicinal Chemistry* **2008**, *16*, 7321-7329; N. Schlienger, T. Beltran, C. Perigaud, I. Lefebvre, A. Pompon, A. M. Aubertin, G. Gosselin, J. L. Imbach, *Bioorganic & Medicinal Chemistry Letters* **1998**, *8*, 3003-3006.
[66] M. Smith, H. G. Khorana, *Journal of the American Chemical Society* **1958**, *80*, 1141-1145.
[67] T. W. Abraham, T. I. Kalman, E. J. McIntee, C. R. Wagner, *Journal of Medicinal Chemistry* **1996**, *39*, 4569-4575.
[68] Mukaiyam.T, Hashimot.M, *Bulletin of the Chemical Society of Japan* **1971**, *44*, 2284-+.
[69] Z. W. Miao, H. Fu, B. Han, Y. Chen, Y. F. Zhao, *Synthetic Communications* **2002**, *32*, 1159-1167; Z.-W. Miao, H. Fu, G.-Z. Tu, Y.-F. Zhao, *Synthetic Communications* **2002**, *32*, 3301-3309.
[70] W. J. Stec, A. Grajkowski, A. Kobylanska, B. Karwowski, M. Koziolkiewicz, K. Misiura, A. Okruszek, A. Wilk, P. Guga, M. Boczkowska, *Journal of the American Chemical Society* **1995**, *117*, 12019-12029.
[71] K. Lu, Q.-P. Duan, L. Ma, D.-X. Zhao, *Bioconjugate Chemistry* **2010**, *21*, 187-202.
[72] N. J. Ede, G. W. Tregear, J. Haralambidis, *Bioconjugate Chemistry* **1994**, *5*, 373-378.
[73] R. Eritja, A. Pons, M. Escarcellar, E. Giralt, F. Albericio, *Tetrahedron* **1991**, *47*, 4113-4120.
[74] L. Chaloin, P. Vidal, P. Lory, J. Méry, N. Lautredou, G. Divita, F. Heitz, *Biochemical and Biophysical Research Communications* **1998**, *243*, 601-608.
[75] J.-P. Bongartz, A.-M. Aubertain, P. G. Milhaud, B. Lebleu, *Nucleic Acids Research* **1994**, *22*, 4681-4688.

[76] N. Venkatesan, B. H. Kim, *Chemical Reviews* **2006**, *106*, 3712-3761; T. S. Zatsepin, D. A. Stetsenko, M. J. Gait, T. S. Oretskaya, *Bioconjugate Chemistry* **2005**, *16*, 471-489; Y. Singh, N. Spinelli, E. Defrancq, *Current Organic Chemistry* **2008**, *12*, 263-290.

[77] P. M. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew Chem Int Ed Engl* **2008**, *47*, 8350-8358.

[78] T. Kubo, Z. Zhelev, B. Rumiana, H. Ohba, K. Doi, M. Fujii, *Organic & Biomolecular Chemistry* **2005**, *3*, 3257-3259.

[79] A. L. Davidson, E. Dassa, C. Orelle, J. Chen, *Microbiology and Molecular Biology Reviews* **2008**, *72*, 317-364.

[80] J. W. Payne, M. W. Smith, *Adv Microb Physiol* **1994**, *36*, 1-80.

[81] M. K. Doeven, J. Kok, B. Poolman, in *Mol Microbiol, Vol. 57*, England, **2005**, pp. 640-649.

[82] J. R. H. Tame, G. N. Murshudov, E. J. Dodson, T. K. Neil, G. G. Dodson, C. F. Higgins, A. J. Wilkinson, *Science* **1994**, *264*, 1578-1581.

[83] C. A. Guyer, D. G. Morgan, J. V. Staros, *J Bacteriol* **1986**, *168*, 775-779.

[84] M. M. Klepsch, M. Kovermann, C. Low, J. Balbach, H. P. Permentier, F. Fusetti, J. W. de Gier, D. J. Slotboom, R. P. Berntsson, in *J Mol Biol, Vol. 414*, 2011 Elsevier Ltd, England, **2011**, pp. 75-85.

[85] T. E. Fickel, C. Gilvarg, *Nat New Biol* **1973**, *241*, 161-163.

[86] S. M. Hammond, A. Claesson, A. M. Jansson, L. G. Larsson, B. G. Pring, C. M. Town, B. Ekstrom, *Nature* **1987**, *327*, 730-732.

[87] M. Benincasa, S. Pacor, R. Gennaro, M. Scocchi, *Antimicrobial Agents and Chemotherapy* **2009**, *53*, 3501-3504.

[88] A. Piffeteau, M. Zamboni, M. Gaudry, *Biochimica Et Biophysica Acta* **1982**, *688*, 29-36.

[89] C. H. Pai, H. C. Lichstein, *Biochim Biophys Acta* **1965**, *100*, 36-42.

[90] P. Hebbeln, D. A. Rodionov, A. Alfandega, T. Eitinger, *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 2909-2914; O. Prakash, Eisenber.Ma, *Journal of Bacteriology* **1974**, *120*, 785-791.

[91] F. Finkenwirth, F. Kirsch, T. Eitinger, *Bioengineered* **2014**, *5*, 129-132.

[92] J. R. Walker, E. Altman, *Applied and Environmental Microbiology* **2005**, *71*, 1850-1855.

[93] G. Gebeyehu, P. Y. Rao, P. SooChan, D. A. Simms, L. Klevan, *Nucleic Acids Research* **1987**, *15*, 4513-4534.

[94] M. A. Kertesz, *Fems Microbiology Reviews* **2000**, *24*, 135-175.

[95] E. Eichhorn, J. R. Van der Ploeg, T. Leisinger, *Journal of Bacteriology* **2000**, *182*, 2687-2695.

[96] M. A. Kertesz, *Research in Microbiology* **2001**, *152*, 279-290.

[97] B. Becker, M. A. Cooper, *Acs Chemical Biology* **2013**, *8*, 105-115.

[98] M. F. Gouzy, M. Lauer, F. Gonzaga, A. Schulz, J. H. Fuhrhop, *Langmuir* **2002**, *18*, 10091-10094; Y. Zhang, J. M. Pelet, D. A. Heller, Y. Dong, D. Chen, Z. Gu, B. J. Joseph, J. Wallas, D. G. Anderson, *Advanced materials (Deerfield Beach, Fla.)* **2013**, *25*, 4641-4645; P. Belmont, A. Aissaoui, M. Hauchecorne, N. Oudrhiri, L. Petit, J.-P. Vigneron, J.-M. Lehn, P. Lehn, *The Journal of Gene Medicine* **2002**, *4*, 517-526.

[99]

[100] A. Ghilardi, D. Pezzoli, M. C. Bellucci, C. Malloggi, A. Negri, A. Sganappa, G. Tedeschi, G. Candiani, A. Volonterio, *Bioconjug Chem* **2013**, *24*, 1928-1936.

[101] N. W. Luedtke, P. Carmichael, Y. Tor, *Journal of the American Chemical Society* **2003**, *125*, 12374-12375.

[102] S. S. Cohen, H. D. Barner, *Proc Natl Acad Sci U S A* **1954**, *40*, 885-893.

[103] R. Then, P. Angehrn, *J Gen Microbiol* **1973**, *76*, 255-263.

[104] C. Atkinson, K. A. Stacey, *Biochim. Biophys. Acta* **1968**, *166*, 705-707.

[105] T. I. Kalman, *Biochem. Biophys. Res. Commun.* **1972**, *49*, 1007-1013.

[106] Freifeld.D, *J. Bacteriol.* **1965**, *90*, 1153-1154; S. I. Ahmad, S. H. Kirk, A. Eisenstark, *Annu Rev Microbiol* **1998**, *52*, 591-625.

[107] J. V. Boyle, M. E. Jones, *J. Bacteriol.* **1970**, *104*, 264-271.

Chapter 2

Tailoring new peptide-nucleotide conjugates (PNCs) for nucleotides delivery in bacterial cells

*This chapter is based on an article published by S. De, E. Groaz and P. Herdewijn, Eur. J. Org. Chem., **2014**, 11, 2322-2348*

ABSTRACT

The design and synthesis of peptide-2'-deoxythymidine-5'-O-monophosphate conjugates as potential active delivery systems of nucleotides in auxotrophic *E. coli* strains is presented. A series of oligopeptides were allowed to react with 5'-O-(dibenzylphosphate)-2'-deoxythymidine or its suitably 3'- derivatized analogues to provide the relevant peptide-nucleotide adducts, owing to the formation of a biolabile chemical connection. Employing strategies based on principles of orthogonal protection and activation, rational variations were made to the linker as well as the peptide moiety in order to tune metabolic stability.

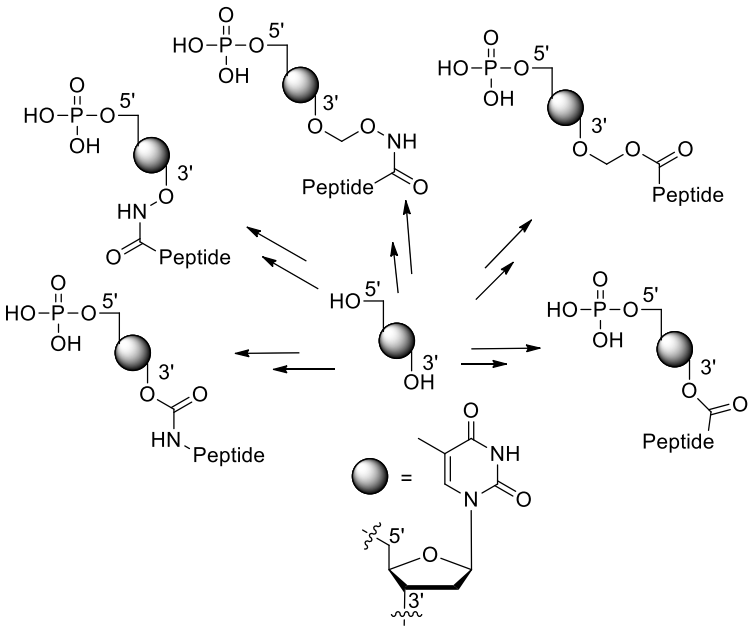
2.1. INTRODUCTION

The ability to encode uncharted heritable information in polymers that are different from DNA and RNA relies upon chemical diversification and enzymatic proliferation of artificial nucleic acids (XNAs).[1, 2] With the aim of controlling and preventing cross contamination, new synthetic genetic sequences will have to be conceived to satisfy a principle of orthogonality with respect to wild biodiversity, leading to the establishment of self-reliant genetic enclaves.[3] Fundamental progress in enzyme evolution has recently been achieved with the discovery of engineered polymerase mutants able to recognize and efficiently incorporate backbone modified nucleotide building blocks (HNA, CeNA, ANA, FANA, TNA, and LNA) from a DNA template. XNA sequences were generated up to 72 nt following canonical base-pairing selectivity and reverse-transcribed into DNA.[4] Nonetheless, major challenges remain to be addressed if XNAs are to reach their true evolutionary potential as genetic reprogramming tools *in vivo*.[5] The molecular assembly of xenopolymers within the host cell requires the intracellular availability of preactivated unnatural precursors, since most modified nucleoside substrates are poorly phosphorylated by host kinases. However, negatively charged nucleotides suffer from low cell-penetrating ability and a known instability as a result of their propensity to undergo hydrolysis to the phosphate-free state of nucleosides by exocytoplasmic phosphatases before being taken up.

Given that, we envisaged that specifically tailored nucleotide conjugates might act as active delivery systems, exploiting membrane uptake systems naturally occurring in bacterial cells for internalization of essential nutrients, to eventually facilitate nucleotide delivery across the cytoplasmatic membrane. In particular, the rationale for our study hinges on the significance of internalization of essential nutrients, to eventually facilitate nucleotide delivery across the cytoplasmatic membrane. In particular, the rationale for our study hinges on the significance of peptide transporters in prokaryotic microorganisms, best represented by the vast family of binding-protein-dependent permeases.[6] The derivatization of ligands with peptides is an attractive concept in medicinal chemistry as it provides cellular access of synthetic antimicrobial compounds, which might be difficult or impossible to be delivered by direct means.[7, 8] Microbial permeases are multicomponent ATP-binding cassette (ABC) proteins mediating active transport of peptides, as source of amino acids, carbon, nitrogen and energy.[9, 10] In Gram-negative species, three classes of peptide permeases (Dpp, Tpp and Opp) have been found, which display a complementary selectivity for different conformations, size and composition of oligopeptide sequences, thus ensuring complete uptake from the peptide pool. Amid the range of periplasmic proteins required for initial molecular recognition and binding by permeases, the oligopeptide binding protein OppA is the most abundant (7-10%) and structurally well-defined.[11, 12] *In vitro* binding studies on peptide complexes of *E. coli* OppA highlighted high affinity and broad substrate specificity for small peptides comprising two up

to five/six L-amino acids, regardless of their sequence.[13, 14] Further investigations have suggested a preference for peptides containing Ala, Gly, Phe[13] and basic residues,[15] in particular lysines, which may result from a negatively charged surface in proximity of the initial binding site.

It has been suggested that the presence of free α-amino and carboxyl groups at the N- and C-terminal respectively were necessary for binding to *E. coli* OppA and subsequent transmembrane translocation into the cell, but protected peptides were also transported although less effectively.[13] However, the presence of a free amino group might trigger extracellular metabolism of the peptide conjugate by peptidases or degradation by chemical means owing to its high reactivity



Linker	Peptide
OCONH	For-M-K-OMe, For-G-K-F-OMe, For-G-βA-F-OMe, For-M-K-F-OMe, For-M-K-A-OMe, For-M-K-K-OMe, For-M-K-A-F-OMe, For-M-K-A-A-F-OMe, A-K-A-OMe, A-K-A
ONHCO	For-M-E-OMe
OCH_2ONHCO	For-M-E-OMe, For-M-E-NH_2
OCH_2OCO	For-G-E-A-OMe
OCO	For-M-K-K, For-G-E-F

Figure 2-1. Chemical diversity in the design of novel peptide-nucleotide conjugates (PNCs) as delivery system of natural and modified nucleotides in prokaryotic cells.

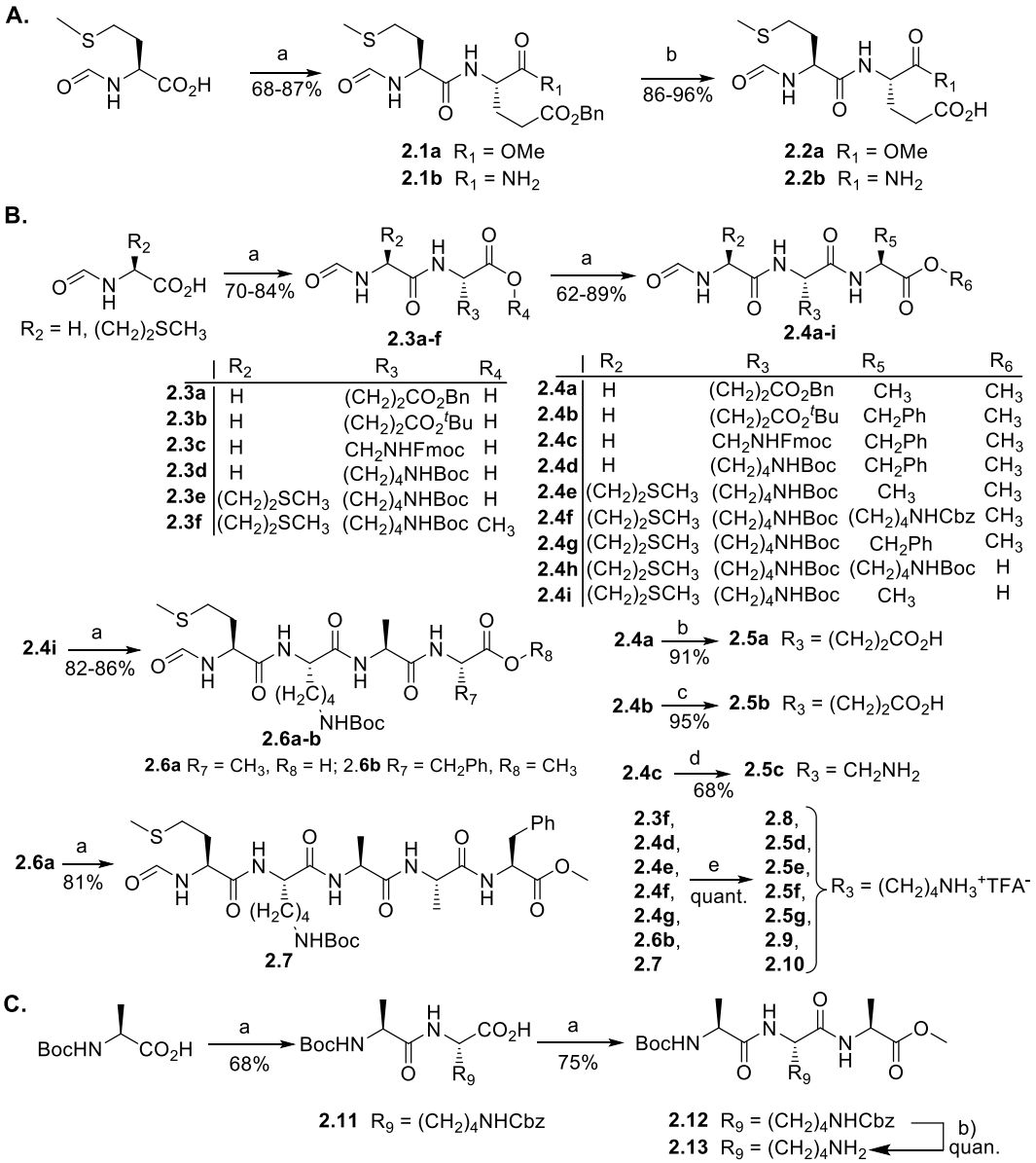
With this knowledge in mind, an approach, which consists of conjugating oligopeptides through a bioreversible linkage to nucleotides, appeared to be logical in attempt to elaborate a new cellular uptake system. It is however anticipated that the success of this proposal would strongly depend on the metabolic stability of the relevant adducts, which ought to be sufficiently stable to be able to cross the cell wall without undergoing premature degradation, while being cleaved at the linker site once in the presence of intracellular enzymes. To assess the viability of our model, we selected 5'-thymidylate (dTMP) nucleotide conjugates as initial synthetic targets, for which reliable nutritional selections are well-established using *E. coli* mutants lacking thymidylate synthase. A conjugate derived from a pyridoxal-pyroglutamyl-tripeptido-nucleotide, which could release the dTMP synthon through an autocatalytic elimination process, was prepared, but purification and scaling-up of this adduct was not straightforward.[16] We describe here a further elaboration of this concept with the synthesis of a large

diversity of peptide-nucleoside conjugate (PNCs) analogues, prepared by covalently linking di- to pentapeptidic linear chain, through various spacers, to the 3'-position on the ribose moiety. Amino acidic residues as well as chemical connections between dTMP and peptides were systematically examined, as summarized in Figure 2-1. The α-amino group at the *N*-terminus was either unmodified or protected as a formylamide, which can be hydrolytically cleaved by *E. coli* peptide deformylase. This enzyme is known to exhibit a preference for substrates containing *N*-formyl methionine, followed by *N*-formyl glycine.[17] The α- or β-carboxyl groups featuring in the second or third residue together with the amino functionality in side chain of Lys were utilized to form the connecting bond. Following internalization of the conjugate in the cell, the peptide might be hydrolysed by *E. coli* PepN, the principal cytosolic aminopeptidase enzyme responsible for the hydrolysis of peptides, whose activity varies in the following order Arg>Ala>Lys>Gly.[18]

2.2. RESULTS AND DISCUSSION

Several strategies have been described in the literature to prepare peptide-oligonucleotide conjugates connected through variously functionalized linkers, such as a thioether, a disulfide, an amide and an oxime moiety etc.[19] For our study, we selected five different linkers (Figure 2-1), which were assumed liable to cleavage inside the bacterial cell, either by a chemical or an enzymatic mechanism, i.e. carbamate (OCONH), ester (OCO), oxyamide (ONHCO), oxymethyleneoxyamide (OCH_2ONHCO), oxymethylene-oxyester (OCH_2OCO). The corresponding peptide-nucleotide conjugates (PNCs) were synthesized by reacting the 3'-hydroxyl group of dTMP with the side chain functionalities of an oligopeptide, for example an amino or a carboxylic acid group.

Firstly, di-, tri-, tetra-, and pentapeptides were conveniently prepared according to the NHS-activated ester method[20] starting from commercially available *N*-Boc protected/formylated amino acids, as shown in Scheme 2-1. In the first step, NHS-activated carboxyester derivatives were formed and subsequently coupled with an incoming amino acid without need of protection at the C-terminal. The side chain functionalities were orthogonally protected as Boc, Fmoc, Cbz, OBn or ^{t}Bu during peptide synthesis and were removed prior to coupling to the free 3'-OH group of thymidine. For the acidic deprotection of the *N*-Boc group, thioanisole was used as radical scavenger,[21] while an excess of Pd/C was used in order to avoid catalyst poisoning during removal of the benzyl group of methionine-containing peptides. In the case of peptide **2.4c**, the Fmoc group was removed using Et_3N in dichloromethane. Ethanol was employed as solvent[22] for *N*-Cbz deprotection of **2.12** to obtain **2.13**, since in the presence of methanol a *N*-methylated side product was formed instead.



Scheme 2-1. Preparation of dipeptides **2.2a**, **2.2b** and **2.8**, tripeptides **2.4h**, **2.5a-g** and **2.13**, tetrapeptide **2.9** and pentapeptide **2.10.** Reagents and conditions: (a) (i) NHS, DCC, THF, 0 °C to r.t., 2-12 h, (ii) Amino acid, $NaHCO_3$, THF/H_2O, 0 °C to r.t., 16 h; (b) 10% Pd/C (Degussa), H_2, EtOH or MeOH, r.t., 6-24 h; (c) TFA, DCM, 0 °C to r.t., 2 h; (d) Et_3N:DCM (2:1), r.t., 72 h; (e) TFA, Thioanisole, DCM, r.t., 4 h.

The construction of PNCs can be accomplished by either linking the peptide at the 3'-position of thymidine followed by 5'-*O*-phosphorylation, or by reversing the order of the reactions sequence. For the synthesis of our first PNC analogue **2.19** featuring a carbamoyl linker (OCONH), it was decided to apply the first proposed protocol, as described in Scheme 2-2. In the first step the 5'-hydroxyl group of 2'-deoxythymidine was selectively protected with a MMTr group[23] leading to 5'-*O*-MMTr-2'-deoxythymidine **2.14,** which was activated at the 3'-OH using carbonyldiimidazole (CDI)[24, 25] or 4-nitrophenyl chloroformate[26] (*via* intermediate **2.15**) and then reacted with the ε-amino group of the

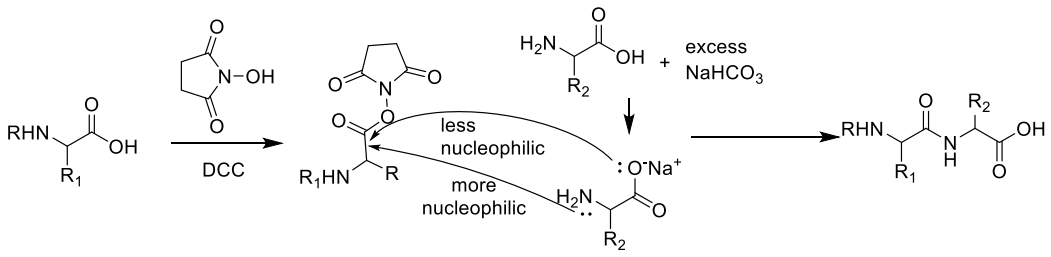
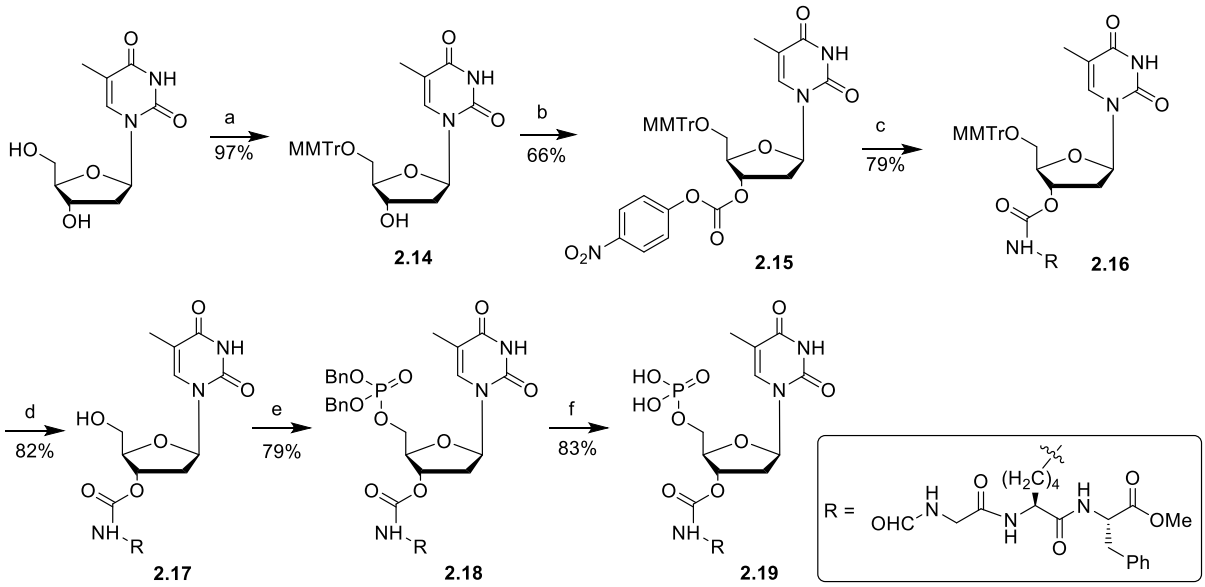


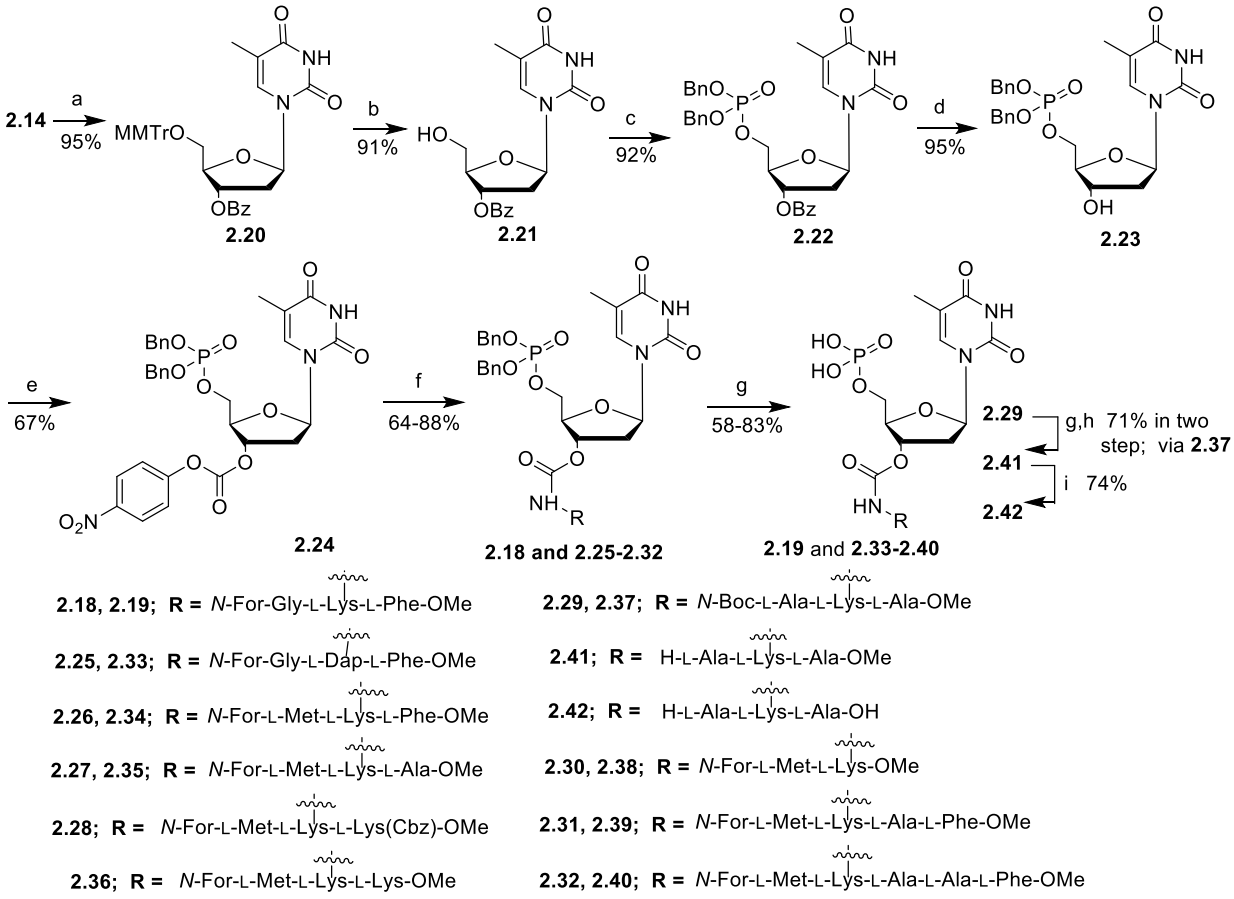
Figure 2-2. Proposed reaction pathway for peptide synthesis without protection of the incoming amino acid.



Scheme 2-2. Synthesis of 3'-*O*-carbamoyl peptide conjugate of dTMP **2.19**. Reagents and conditions: (a) MMTrCl, Et_3N, DMAP, DMF, r.t., 4 h; (b) 4-nitro phenyl chloroformate, Py, DCM, 0 °C to r.t., 24 h; (c) Peptide **2.5d**, Et_3N, DCM, r.t.; (d) 80% AcOH, r.t., 3.5 h; (e) (i) Dibenzyl *N,N*-diisopropylphosphoramidite, 0.45M tetrazole in ACN, DCM, r.t., 12 h, (ii) H_2O_2, -40 °C to r.t., 2 h; (f) 10% Pd/C (Degussa), H_2, Et_3N, MeOH.

lysine residue of peptide **2.5d** to afford compound **2.16**. After detritylation, the phosphate group was introduced at the 5'-position using dibenzyl-*N,N*-diisopropyl phosphoramidite and hydrogen peroxide as oxidizing agent. The resulting thymidine-5' dibenzyl phosphate product **2.18** was subjected to hydrogenation in the presence of Pd/C to produce the free phosphate functionality. However, this strategy was not applicable to the synthesis of *N*-formyl-L-methionine-containing analogues presumably due to the instability of the thiomethyl group in the course of the oxidation of $P^{(III)}$ to $P^{(V)}$, even when milder oxidizing reagents, such as a diluted solution of I_2 in pyridine-THF-H_2O, were used.[27, 28] Therefore, the route towards compound **2.19** was repeated using as starting material 3'-benzoyl thymidine **2.21**, as described in Scheme 2-3. The dibenzylphosphate group at the 5'-position

of **2.21** was introduced first, followed by peptide conjugation. The same synthetic method was extended to afford nine different PNCs **2.18** and **2.25-2.32** built upon a carbamoyl linker.

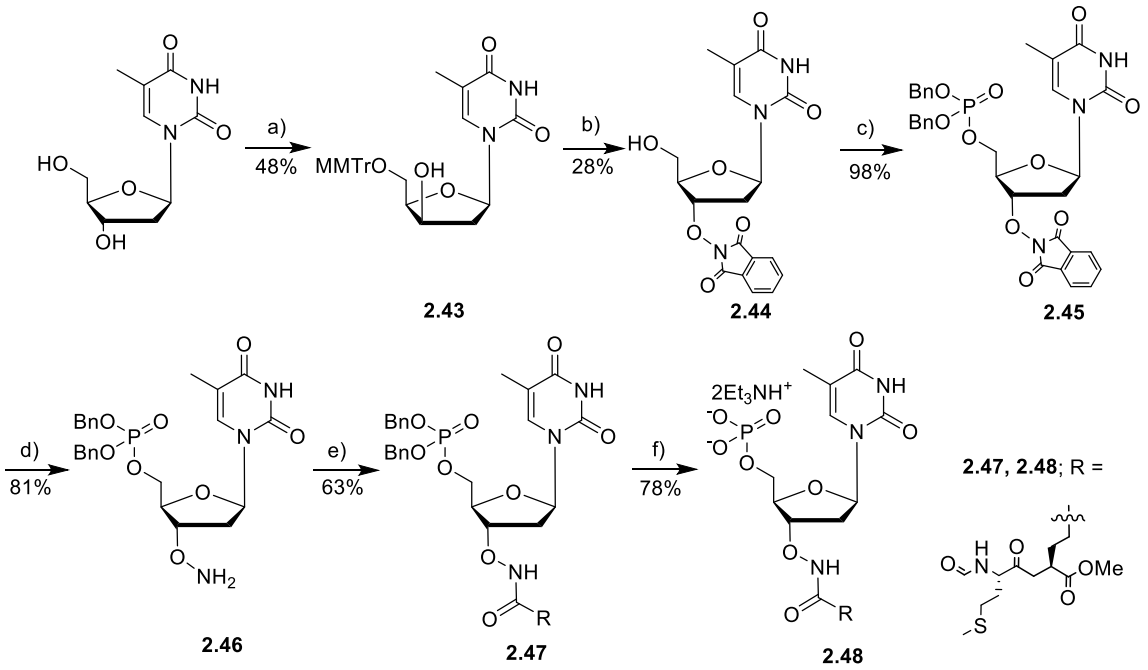


Scheme 2-3. Synthesis of 3'-carbamoyl peptide conjugates of dTMP **2.19** and **2.33-2.42**. Reagents and conditions: (a) BzCl, Py, 0 °C, 3 h; (b) 80% AcOH, r.t., 3.5 h; (c) (i) Dibenzyl *N,N*-diisopropylphosphoramidite, 0.45M tetrazole in ACN, DCM, r.t., 12 h, (ii) H_2O_2, -40 °C to r.t., 2 h; (d) 7N NH_3 in MeOH, 0 °C to r.t., 24 h; (e) 4-nitro phenyl chloroformate, Py, DCM, 0 °C to r.t., 24 h; (f) Peptide-NH_2 (**2.5c-2.5g**, **2.8-2.10** or **2.13**), Et_3N, DCM, r.t.; (g) 10% Pd/C (Degussa), H_2, MeOH or EtOH, r.t., 24 h; (h) TFA, Thioanisole, H_2O, r.t., 3 h; (i) LiOH, THF:MeOH:H_2O (1:1:1), r.t., 3 h.

With regard to this second route, protection of the secondary OH group was necessary to obtain the corresponding phosphorylated nucleoside **2.23** in good yield.[29, 30, 31] Thus, we elected to protect the free hydroxyl functionality of compound **2.14** using benzoyl chloride[32]. Standard MMTr deprotection proceeded as expected using 80% AcOH to give 3'-*O*-benzoyl-2'-deoxythymidine **2.21** which was in turn phosphorylated with dibenzyl *N,N*-diisopropyl-phosphoramidite and 1*H*-tetrazole[33] at the 5'-position and then oxidized in the presence of hydrogen peroxide to furnish 5'-dibenzylphosphate-3'-*O*-benzoyl-2'-deoxythymidine **2.22** in 92% yield over two-steps. Debenzoylation was affected by treatment of **2.22** with saturated methanolic ammonia to provide key intermediate **2.23** in 95% yield, which served as a common substrate for the subsequent coupling step

exploring various 3'-*O* linkers. In Scheme 2-3, the 3'-OH was derivatized with 4-nitrophenyl chloroformate in pyridine to furnish 4-nitrophenyl carbonate **2.24** in moderate yield.[34] The activated carbonate was then treated with amino- peptides **2.5c-2.5g**, **2.8-2.10** and **2.13** to generate carbamates **2.18** and **2.25-2.32** with yields ranging from 64 to 88%, depending on the peptide used. Finally, hydrogenolysis of the benzyl groups was carried out under mild conditions, in order to prevent reduction of the double bond of the thymine ring, using 10% palladium on carbon (Degussa) or 20% $Pd(OH)_2/C$ at atmospheric pressure, smoothly affording phosphate derivatives **2.19** and **2.33-2.40** in 58-83% yields.

Notably, debenzylation by hydrogenation was successful for conjugates containing methionine in the peptide chain. A larger excess of Pd/C was employed during the reaction to avoid catalyst poisoning deriving from the thiomethyl group. In most of the cases, except for compounds **2.36** where a free amino group is present in the peptidic side chain, 2 equivalents of Et_3N were employed to neutralize the strong acidity of the phosphoric acid group during *O*-dibenzyl deprotection. The terminal amino group of **2.37** was deprotected with TFA affording conjugate **2.41**, which was then treated with LiOH to liberate the terminal carboxyl group leading to compound **2.42**.



Scheme 2-4. Synthesis of 3'-oxyamide peptide conjugate of dTMP **2.48**. Reagents and conditions: (a) (i) MMTrCl, Et_3N, Py, r.t., 16 h, (ii) MsCl, Et_3N, r.t., 2 h, (iii) NaOH, EtOH, reflux, 1.5 h; (b) (i) *N*-hydroxy phthalimide, PPh_3, DIAD, Toluene, 0 °C to r.t., 2 h, (ii) 80% AcOH, r.t., 3 h; (c) (i) Dibenzyl *N*,*N*-diisopropylphosphoramidite, 0.45M tetrazole in ACN, DCM, r.t., 12 h, (ii) H_2O_2, -40 °C to r.t., 2 h; (d) 4% $MeNH_2$, EtOH, r.t., 0.5 h; (e) Dipeptide **2.2a**, DCC, DMAP, DCM/DMF, r.t., 24 h; (f) 10% Pd/C (Degussa), H_2, MeOH, Et_3N, r.t., 24 h.

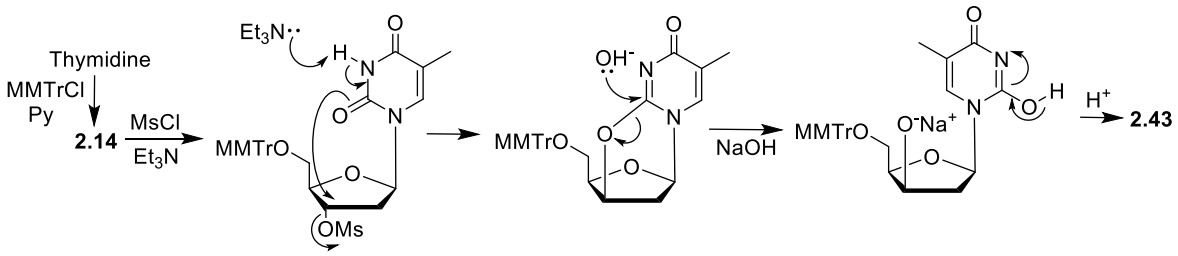


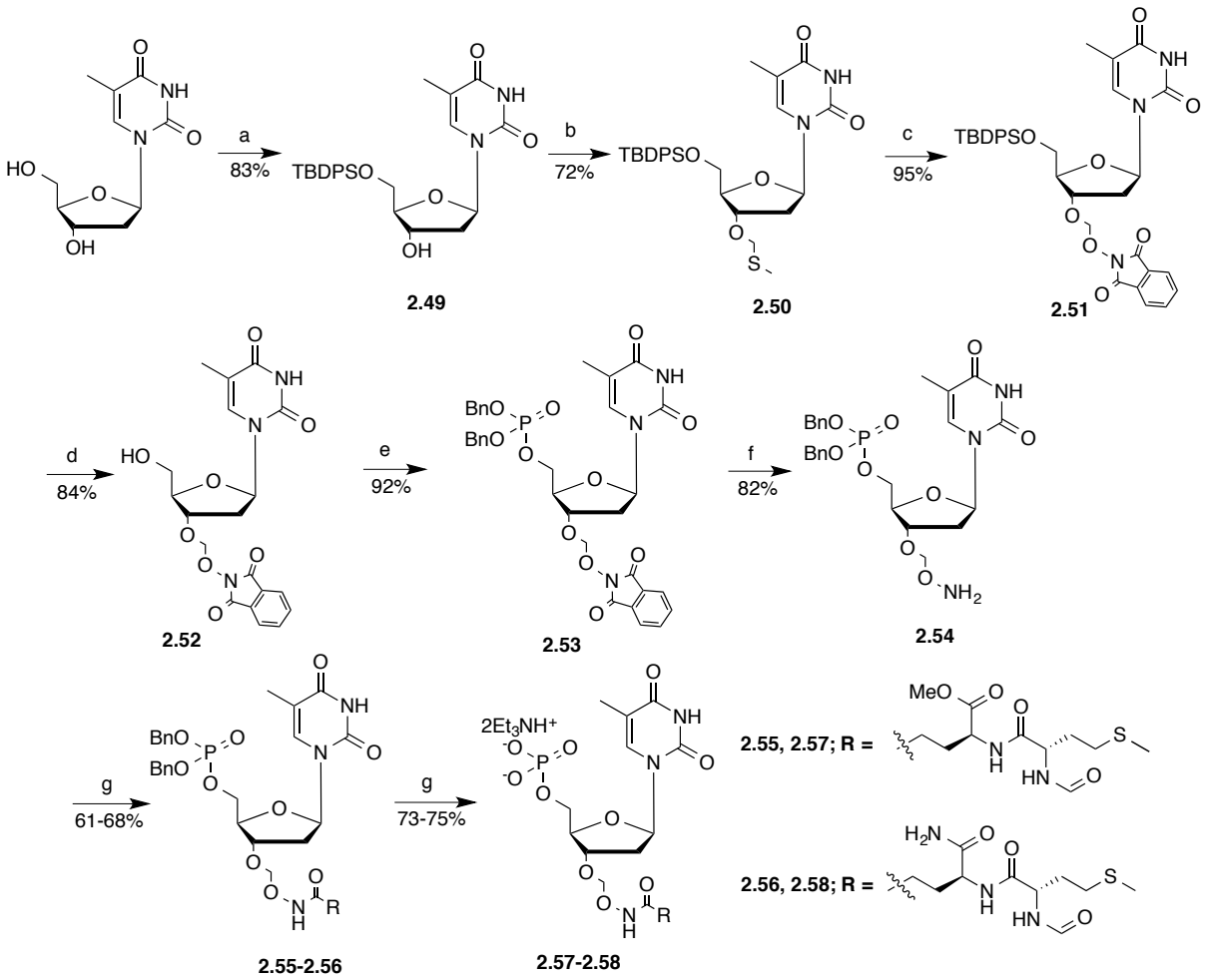
Fig 2-3: Proposed reaction mechanism for the formation of **2.43** from thymidine via 2-3'-anhydro intermediate formation.

The synthesis of the 3'-oxyamide linked PNC **2.48** is described in Scheme 2-4. The 3'-oxyamine intermediate **2.46** was synthesized starting from 2'-deoxythymidine according to modified literature procedures.[35, 36, 37] 5'-*O*-MMTr-2'-deoxy-*xylo*-thymidine **2.43** was obtained from 2'-deoxythymidine in 48% yield over three steps.[38]

The protected hydroxylamine functionality was formed *via* Mitsunobu inversion using *N*-hydroxyphthalimide. Removal of MMTr and phosphorylation at the 5'-position, as previously described, yielded the protected hydroxylamine **2.45**, which was then treated with 4% methylamine to restore the 3'-ONH_2 group. Although stronger nucleophilic reagents like hydrazine hydrate or hydroxylamine are more efficient for phthalimide deprotection, our reagent of choice was diluted methylamine, to minimize rate of side reactions at the protected 5'-phosphate moiety.

Coupling between the 3'-oxyamino group of compound **2.46** with the free carboxylic acid group of dipeptide **2.2a** in the presence of DCC, led to benzyl-protected oxyamide **2.47**, which was subjected to catalytic hydrogenation using Pd/C and Et_3N at atmospheric pressure. Final purification by RP-HPLC yielded **2.48** as a triethylammonium salt. The common key intermediate towards 3'-oxymethyleneoxyamide PNCs **2.57** and **2.58** is compound **2.54** (Scheme 2-5), which can be coupled with different peptides exhibiting a free carboxylic acid group in their side chain residue.[39, 40] Product **2.54** was obtained starting from 5'-*O*-protected-3'-*O*-methylenethiomethyl-2'-deoxythymidine **2.50**, which was synthesized modifying a procedure from the literature.[41] The acid stable TBDPS group was selected to mask the 5'-position in view of potential complications in a later stage of the synthetic scheme. The 3'-thioacetal group of **2.50** (obtained via Pummerer rearrangement from **2.49**, fig. 2-4) was converted by treatment with sulfuryl chloride into its chloromethylether derivative, which, without further purification, was reacted with *N*-hydroxyphthalimide to obtain 3'-*O*-aminooxymetylene derivative **2.51**. The 5'-TBDPS protection was removed by action of $Et_3N{\cdot}3HF$ and the free 5'-hydroxyl group was phosphorylated using our general protocol. Phthalimide deprotection followed by DCC coupling with dipeptides **2.2a** and **2.2b** yielded the conjugates **2.55** and **2.56** respectively. Benzyl deprotection and HPLC purification as previously described, yielded

the 5'-*O*-monophosphate of 2'-deoxythymidine-3'-oxymethyleneoxyamide peptide conjugates **2.57** and **2.58** as triethylammonium salts.



Scheme 2-5. Synthesis of 3'-oxymethyleneoxyamide peptide conjugates of dTMP **2.57** and **2.58**. Reagents and conditions: (a) (i) TBDPS-Cl, Imidazole, DMF, r.t., 4 h; (b) DMSO, Ac_2O, AcOH, r.t., 48 h; (c) (i) SO_2Cl_2, DCM, 0 °C to r.t., 2 h, (ii) *N*-Hydroxy phthalimide, DBU, DCM, r.t., 24 h; (d) $Et_3NH{\cdot}3HF$, THF, r.t., 36 h; (e) (i) Dibenzyl *N*,*N*-diisopropylphosphoramidite, 0.45M tetrazole in ACN, DCM, r.t., 12 h, (ii) H_2O_2, -40 °C to r.t., 2 h; (f) 4% $MeNH_2$, EtOH, r.t., 0.5 h; (g) Dipeptide **2.2a** or **2.2b**, DCC, DMAP, DCM/DMF, r.t., 24h; (h)10% Pd/C (Degussa), H_2, Et_3N, MeOH, r.t., 24 h.

Due to its propensity to undergo cleavage by cellular carboxyesterase, the oxymethylene ester moiety has been described in the literature in the generation of bioavailable antiretroviral prodrugs based on acyclic nucleosides,[42, 43] and as biolabile 2'-*O*-protecting group for the development of short synthetic double-stranded RNA sequences with therapeutic applications *in vivo*.[44, 45, 46] The formation of 3'-*O*-methylacyloxy PNC **2.62** essentially followed known procedures for similar

transformations, by activating previously synthesized intermediate **2.50** and using the β-carboxyl group of tripeptide **2.5a** as a nucleophile (Scheme 2-6).

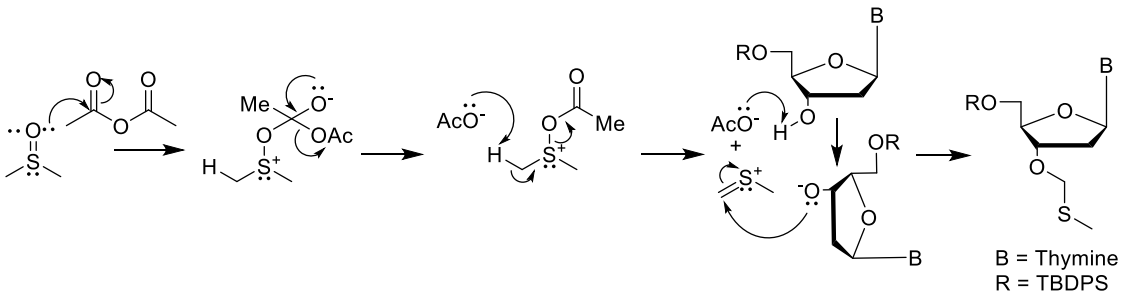
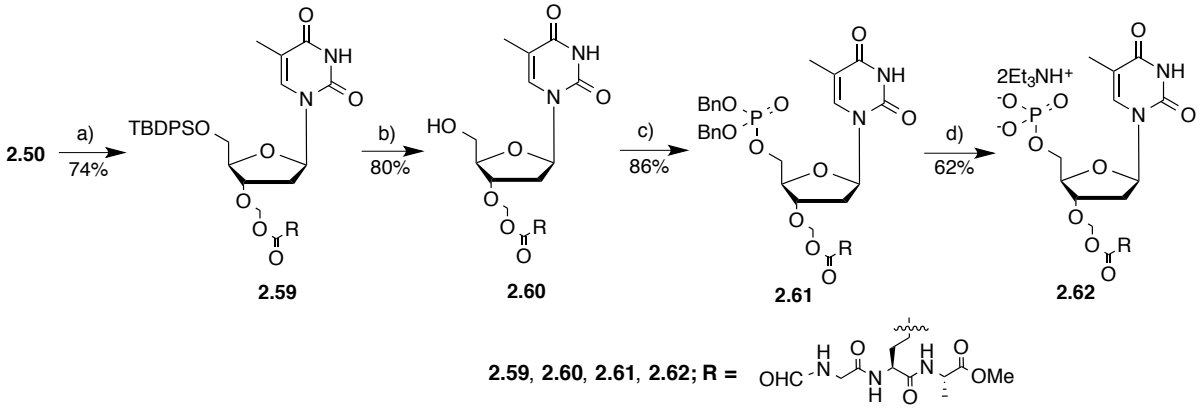


Fig 2-4: Proposed reaction mechanism for the formation of **2.50** from **2.49** via Pummerer rearrangement.

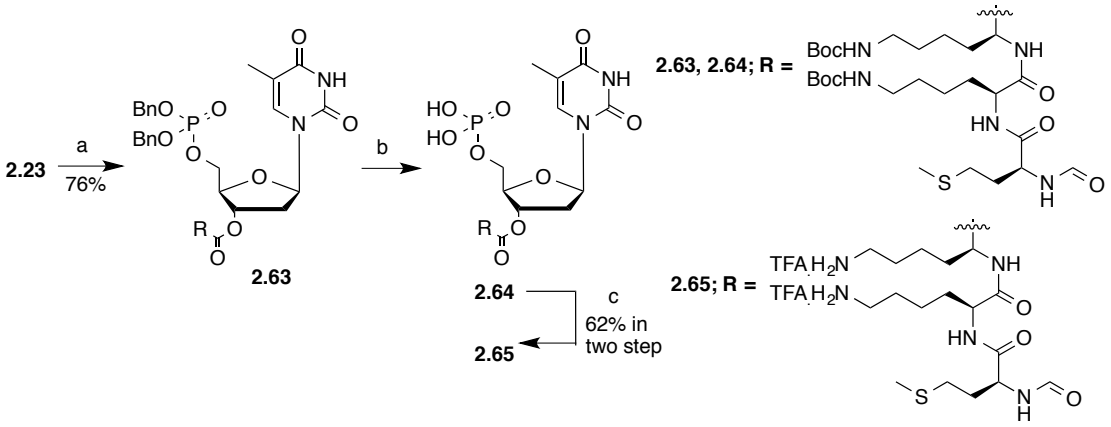
A first attempt using NIS/NCS as activating agent[47] did not give satisfactory yields of compound **2.59**, therefore we switched more successfully to the sulfuryl chloride method.[48, 49] TBDPS deprotection using $Et_3N{\cdot}3HF$ and phosphorylation at the 5'-position was carried out as described in Scheme 2-5, yielding compound **2.61**. Finally, $Pd(OH)_2$-mediated hydrogenolysis in the presence of Et_3N furnished triethylammonium salt **2.62** in moderate yield, due to undesirable side reactions involving cleavage of the linker, giving rise to dTMP and peptide as by-products.



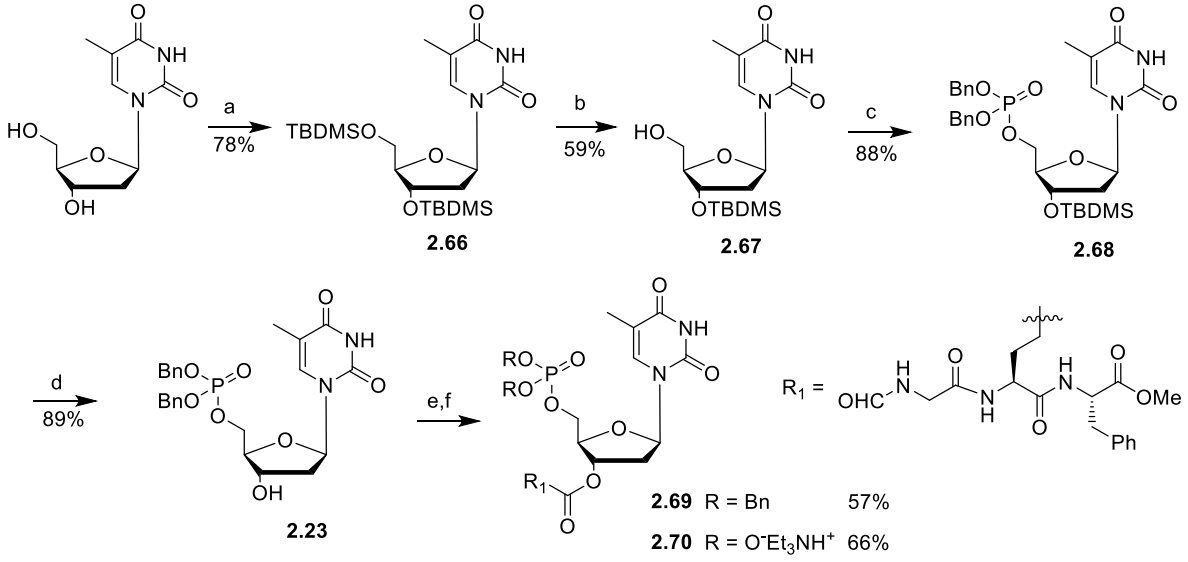
Scheme 2-6. Synthesis of 3'-oxymethyleneoxyester peptide conjugate of dTMP **2.62**. Reagents and conditions: (a) (i) SO_2Cl_2, DCM, 0 °C to r.t., 2 h, (ii) Peptide-CO_2H **2.5a**, DBU, DCM, r.t., 24 h; (b) $Et_3NH{\cdot}3HF$, THF, r.t., 36 h; (c) (i) Dibenzyl *N,N*-diisopropylphosphoramidite, 0.45M tetrazole in ACN, DCM, r.t., 12 h, (ii) H_2O_2, -40 °C to r.t., 2 h; (e) 20% $Pd(OH)_2/C$, H_2, $NaHCO_3$, $EtOH/H_2O$, r.t., 1.5 h.

The synthesis of PNCs with an ester linker was thought to be particularly promising in view of the known susceptibility of such functional group to intracellular hydrolysis by bacterial esterases. Analogue **2.65** was obtained from previously prepared protected-2'-deoxythymidine-5'-*O*-monophosphate **2.23** by standard coupling to the free terminal carboxylate of tripeptide **2.4h** in the

presence of DCC and DMAP as catalyst,[50] in good yield. Deprotection of the dibenzyl phosphate *via* catalytic hydrogenation led to **2.64**, which, without further purification, underwent removal of the NH-Boc protecting groups in the presence of TFA to furnish final compound **2.65** as a TFA salt in 62% yield over two steps (Scheme 2-7). When Cbz was used in place of Boc, attempts to remove all protecting groups in one step using catalytic hydrogenation resulted in a complex mixture of degradation products.



Scheme 2-7. Synthesis of 3'-carboxyester peptide conjugate of dTMP **2.65**. Reagents and conditions: (a) Peptide-CO_2H **2.4h**, DCC, DMAP, DCM:DMF (5:1), r.t., 24 h; (b) 10% Pd/C (Degussa), H_2, MeOH, r.t., 24 h; (c) TFA, Thioanisole, H_2O, r.t., 5 h.



Scheme 2-8. Synthesis of 3'-carboxyester peptide conjugate of dTMP **2.70**. Reagents and conditions: (a) TBDMSCl, Im, dry DMF; (b) TFA:H_2O (10:1), DCM; (c) (i) Dibenzyl *N*,*N*-diisopropylphosphoramidite, 0.45M tetrazole in ACN, DCM, 0 °C to r.t., 12 h, (ii) H_2O_2, -25 °C to r.t., 2 h; (d) $Et_3N{\cdot}3HF$, THF, 24 h; (e) Peptide-CO_2H **2.5b**, DCC, DMAP, DMF, DCM, r.t., 24 h; (b) 10% Pd/C (Degussa), H_2, THF, r.t., 24 h.

The key intermediate 5'-*O*-(dibenzylphosphate)-2'-deoxythymidine **2.23** could also be prepared through an alternative sequence of protection/deprotection steps. Hence, the synthesis, illustrated in Scheme 2-8, started with the bis-silylation of thymidine to form intermediate **2.66**, which was selectively deprotected at the 5'-position using a 10:1 ratio of TFA in H_2O in good yield (59%).[51] Compound **2.67** was then subjected to standard phosphorylation, followed by TBDMS cleavage, providing the required substrate **2.23** with which to carry out peptide coupling. In this step the use of a milder reagent such as $Et_3N{\cdot}HF$ was found to afford higher chemical yields than TBAF. In the presence of the free β-carboxyl group of tripeptide **2.5b**, DCC and DMAP, **2.23** was cleanly converted into conjugate **2.69** and eventually its triethylammonium salt **2.70**.

2.3. CONCLUSION

The derivatization of ligands with peptides can be used to achieve a large improvement in the cellular uptake of hydrophilic structures. We have reported the application of this methodology to nucleotides, which has led to the development of a broad variety of peptide-5'-thymidylate conjugates featuring linker moieties such as carbamate, ester, oxyamide, oxymethyleneoxyamide and oxymethyleneoxyester, all potentially liable to enzymatic or chemical cleavage inside the bacterial cell. Final deprotection of all structures was achieved in high yield by palladium-catalyzed hydrogenolysis. Of particular note is the unexpected tolerance of the thiomethyl functionality present in methionine-containing PNCs to the debenzylation conditions.

2.4. EXPERIMENTAL SECTION

General information

For all reactions, analytical grade solvents were used. All moisture-sensitive reactions were carried out under an argon or nitrogen atmosphere in oven-dried glassware (135 ºC). Reaction temperatures are reported as bath temperatures. Precoated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ = 254 nm). Products were purified by flash chromatography on ICN silica gel 63-200, 60 Å. 1H, 13C and 31P NMR spectra were recorded on Bruker Avance 300 MHz, 500 MHz or 600 MHz spectrometers. For final compounds, 1H and 13C resonance assignments were made using 2D NMR correlation experiments (COSY, gHSQC and gHMBC). For sake of clarity, NMR signals of protons and carbons for sugar and base moieties are indicated with and without a prime, respectively. Chemical shifts were referenced to residual solvent signals at δ H/C 7.26/77.00 ($CDCl_3$), 3.31/49.10 (MeOD) and 2.50/39.50 (DMSO-*d6*) relative to TMS as internal standard. Coupling constants are expressed in hertz (Hz). Splitting patterns are reported as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet) and br (broad). High resolution mass spectra were acquired on a quadrupole orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 μL/min and spectra were obtained in positive or negative ionization mode with a resolution of 15000 (FWHM), using leucine enkephalin as lock mass. All the methods use $MeCN/H_2O$ gradients. Water contains either 0.1 % TFA

or 0.1 % NH_3. All final compounds were purified by preparative RP-HPLC (Xbridge™ Prep C18 5µm OBD 19 x 150 mm column).

General procedures for oligopeptides synthesis. Method (a). DCC (1.3 eq.) was added to a solution of *N*-Boc- or *N*-formyl-amino acid (1 eq.) and *N*-hydroxysuccinimide (1.2 eq.) in dry THF (~4 mL/mmol) at 0 °C, and the mixture was stirred for an additional 0.5 h at 0 °C, then for 2-6 h (monitoring by TLC) at r.t. under an argon atmosphere. The resulting DCU was filtered off, and the filtrate was concentrated *in vacuo* to obtain the corresponding activated ester. This compound was then dissolved in dry THF (2 mL/mmol), and to this solution was then added a suspension of a second amino acid (1.2 eq.) and $NaHCO_3$ (4 eq.) in water (2 mL/mmol) at 0 °C under an argon atmosphere. The reaction mixture was then stirred at r.t. for 16 h, cooled and neutralized by addition of 1N HCl. The aqueous layer was extracted 3 times with ethyl acetate, the combined organic fractions were dried over Na_2SO_4, filtered and concentrated under reduced pressure. The resulting crude residue was purified by column chromatography on silica gel (gradient: CH_2Cl_2/MeOH 98:2, v/v; 96:4, v/v; 94:6, v/v) to obtain the desired compound. **Method (b)**. A minor modification of Method (a) was applied during the work-up stage when the carboxyl group at the C-terminal in the peptides was protected as ester or amide. Thus, after amide coupling with the activated ester, the reaction mixture was diluted with 10% $NaHCO_3$ and extracted with ethyl acetate to ease the purification step by removing the acidic by-products.

***N*-For-L-Met-L-Glu(OBn)OMe (2.1a):** Following Method (b), dipeptide **2.1a** was obtained as a white solid (1.01 g, 87%), in two steps, starting from *N*-For-L-Met-OH (0.5 g, 2.82 mmol), *N*-hydroxysuccinimide (0.39 g, 3.38 mmol) and DCC (0.76 g, 3.67 mmol) in THF (20 mL), and H-L-Glu(OBn)OMe·TFA salt (1.18 g, 3.24 mmol, prepared from Boc-L-Glu(OBn)OMe and TFA in DCM with 1 eq. of thioanisole), $NaHCO_3$ (0.95 g, 11.3 mmol) in 1:1 THF/H_2O (30 mL) ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.16 (s, 1H, C*H*O), 7.35-7.32 (m, 5H, Ar-*H*), 7.03 (d, *J* = 7.7 Hz, 1H, N*H*-Glu), 6.47 (d, *J* = 7.5 Hz, 1H, N*H*-Met), 5.12 (s, 2H, OC*H*$_2$Ph), 4.77-4.70 (m, 1H, αC*H*-Met), 4.64-4.59 (m, 1H, αC*H*-Glu), 3.74 (s, 3H, OC*H*$_3$), 2.64-2.55 (m, 2H, γC*H*$_2$-Met), 2.52-2.43 (m, 2H, γC*H*$_2$-Glu), 2.29-1.96 (m unresolved, 7H, βC*H*$_2$-Glu, SC*H*$_3$ and βC*H*$_2$-Met); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.7 (δCO-Glu), 171.8 (αCO-Glu), 170.7 (CO-Met), 160.9 (CHO), 135.8 (1C-Ph), 128.7 (Ar-C), 128.5 (Ar-C), 128.4 (Ar-C), 66.8 (OCH_2Ph), 52.7 (OCH_3), 52.0 (αC-Glu), 50.7 (αC-Met), 31.8 (βC-Met), 30.4 (γC-Glu), 29.9 (γC-Met), 27.0 (βC-Glu), 15.2 (SCH_3); HRMS for $C_{19}H_{26}N_2O_6S$ $[M-H]^-$ calcd.: 411.1584, found: 411.1584.

***N*-For-L-Met-L-Glu(OBn)-NH_2 (2.1b):** Following Method (b), dipeptide **2.1b** was obtained as a white solid (0.76 g, 68%), in two steps, starting from *N*-For-L-Met-OH (0.5 g, 2.82 mmol), *N*-hydroxysuccinimide (0.39 g, 3.38 mmol) and DCC (0.76 g, 3.67 mmol) in THF (20 mL), and H-L-Glu(OBn)-NH_2·TFA salt (1.09 g, 3.24 mmol, prepared from Boc-L-Glu(OBn)-NH_2 and TFA in DCM with 1 eq. of thioanisole), $NaHCO_3$ (0.95 g, 11.3 mmol) in 1:1 THF/H_2O (30 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.32 (d, *J* = 8.2 Hz, 1H, N*H*-Met), 8.05 (d, *J* = 8.0 Hz, 1H, N*H*-Glu), 8.02 (s, 1H, C*H*O), 7.38-7.34 (m, 5H, Ar-*H*), 7.30 (br s, 1H, N*H*-$CONH_2$), 7.09 (br s, 1H, N*H*-$CONH_2$), 5.08 (s, 2H, OC*H*$_2$Ph), 4.43-4.36 (m, 1H, αC*H*-Met), 4.24-4.16 (m, 1H, αC*H*-Glu), 2.44 (t, *J* = 7.9 Hz, 2H, γC*H*$_2$-Met), 2.36 (t, *J* = 7.8 Hz, 2H, γC*H*$_2$-Glu), 2.02 (s, 3H, SC*H*$_3$), 1.98-1.86 (m, 2H, βC*H*$_2$-Glu), 1.81-1.71 (m, 2H, βC*H*$_2$-Met); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.7 (αCO-Glu), 172.1 (δCO-Glu), 170.7 (CO-Met), 161.1 (CHO), 136.2 (1C-Ph), 128.4 (Ar-C), 128.0 (Ar-C), 127.9 (Ar-C), 65.5 (OCH_2Ph), 51.6 (αC-Glu), 50.6 (αC-Met), 31.9 (βC-Met), 30.1 (γC-Glu), 29.3 (γC-Met), 27.1 (βC-Glu), 14.6 (SCH_3); HRMS for $C_{18}H_{25}N_3O_5S$ $[M+Na]^+$ calcd.: 418.1407, found: 418.1407.

***N*-For-Gly-L-Glu(OBn)OH (2.3a):** Following Method (a), dipeptide **2.3a** was obtained as a white solid (2.47 g, 79%) in two steps, starting from *N*-For-Gly-OH (1.0 g, 9.70 mmol), *N*-hydroxysuccinimide (1.34 g, 11.6 mmol), DCC (2.6 g, 12.6 mmol) in THF (30 mL) and H-L-Glu(OBn)-OH·TFA salt (3.92 g, 11.2 mmol, prepared from Boc-L-Glu(OBn)-OH and TFA in DCM with 1 eq. of thioanisole), $NaHCO_3$ (3.26 g, 38.8 mmol) in 1:1 THF/H_2O (60 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.73 (br s, 1H, CO_2H), 8.23-8.21 (m, 2H, 2 x N*H*), 8.07 (s, 1H, C*H*O), 7.37-7.34 (m, 5H, Ar-*H*), 5.09 (s, 2H, OCH_2Ph), 4.30-4.23 (m, 1H, αC*H*-Glu), 3.78 (app t, *J* = 5.3 Hz, 2H, αCH_2-Gly), 2.43 (t, *J* = 7.7 Hz, 2H, γCH_2-Glu), 2.10-1.78 (m, 2H, βCH_2-Glu), 1.81-1.71 (m, 2H, βCH_2-Met); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 173.0 (αCO-Glu), 172.1 (δCO-Glu), 168.5 (CO-Gly), 161.4 (CHO), 136.2 (1C-Ph), 128.4 (Ar-C), 128.0 (Ar-C), 127.9 (Ar-C), 65.5 (OCH_2Ph), 51.1 (αC-Glu), 40.3 (αC-Gly), 30.0 (γC-Glu), 26.3 (βC-Glu); HRMS for $C_{15}H_{18}N_2O_6$ $[M-H]^-$ calcd.: 321.1092, found: 321.1090.

***N*-For-Gly-L-Glu(O^{t}Bu)OH (2.3b):** Following Method (a), dipeptide **2.3b** was obtained as a white solid (2.79 g, 84%), in two steps, starting from *N*-For-L-Gly-OH (1.00 g, 9.70 mmol), *N*-hydroxysuccinimide (1.34 g, 11.6 mmol) and DCC (2.60 g, 12.6 mmol) in THF (16 mL), and H-L-Glu(O^{t}Bu)-OH (2.46 g, 12.1 mmol), $NaHCO_3$ (2.24 g, 26.7 mmol, 2.2 eq.) in 1:1 THF/H_2O (30 mL). ^{1}H NMR (300 MHz, MeOD) δ = 8.18 (s, 1H, C*H*O), 4.49-4.45 (m, 1H, αC*H*-Glu), 3.99 (br s, 2H, αCH_2-Gly), 2.34 (t, *J* = 7.5 Hz, 2H, γCH_2-Glu), 2.21-2.09 (m, 1H, βCH_2-Glu), 1.98-1.88 (m, 1H, βCH_2-Glu), 1.43 (s, 9H, C(CH_3)$_3$); ^{13}C NMR (75 MHz, MeOD) δ = 173.0 (CO-Glu), 172.1 (CO-Glu), 169.3 (CO-Gly), 162.6 (CHO), 80.2 (1C-^{t}Bu), 51.2 (αC-Glu), 40.2 (αC-Gly), 30.8 (γC-Glu), 26.7 (^{t}Bu), 26.2 (βC-Glu); HRMS for $C_{12}H_{19}N_2O_6$ $[M-H]^-$ calcd.: 287.1248, found: 287.1245.

***N*-For-Gly-L-β-amino-Ala(Fmoc)OH (2.3c):** Following Method (a), dipeptide **2.3c** was obtained as a white solid (1.54 g, 70.3%) in two steps, starting from *N*-For-Gly-OH (0.55 g, 5.33 mmol), *N*-hydroxysuccinimide (0.74 g, 6.40 mmol), DCC (1.43 g, 6.93 mmol) in THF (20 mL) and H-L-β-amino-Ala(Fmoc)-OH·TFA salt (2.70 g, 6.13 mmol, prepared from Boc-L-β-amino-Ala(Fmoc)-OH and TFA in DCM with 1 eq. of thioanisole), $NaHCO_3$ (1.79 g, 21.4 mmol) in 1:1 THF/H_2O (50 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.19 (br s, 1H, CO_2H), 8.24 (t, *J* = 5.0 Hz, 1H, N*H*), 8.17 (d, *J* = 7.9 Hz, 1H, N*H*), 8.08 (s, 1H, C*H*O), 7.89 (d, *J* = 7.4 Hz, 2H, Ar*H*-Fmoc), 7.69 (d, *J* = 7.3 Hz, 2H, Ar*H*-Fmoc), 7.44-7.31 (m, 4H, Ar*H*-Fmoc), 4.35-4.22 (unresolved m, 4H, *αH*-β-amino-Ala, OCH_2-Fmoc and C*H*-Fmoc), 3.80 (d, *J* = 5.8 Hz, 2H, αCH_2-Gly), 3.50-3.30 (m, 2H, βCH_2-β-amino-Ala merged with DMSO); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 171.5 (CO-β-amino-Ala), 168.5 (CO-Gly), 161.4 (CHO), 156.3 (OCONH), 143.8 (Ar-C), 140.7 (Ar-C), 127.6 (Ar-C), 127.1 (Ar-C), 125.2 (Ar-C), 120.1 (Ar-C), 65.6 (OCH_2-Fmoc), 52.3 (αC-β-amino-Ala), 46.6 (CH-Fmoc), 41.5 (αC-Gly), 30.7 (βC-β-amino-Ala); HRMS for $C_{21}H_{21}N_3O_6$ $[M-H]^-$ calcd.: 410.1357, found: 410.1357.

***N*-For-Gly-Lys(Boc)OH (2.3d):** Following Method (a), dipeptide **2.3d** was obtained as a white solid (2.60 g, 81%) in two steps, starting from *N*-For-Gly-OH (1.0 g, 9.70 mmol), *N*-hydroxysuccinimide (1.34 g, 11.6 mmol), DCC (2.60 g, 12.6 mmol) in THF (35 mL) and H-L-Lys(Boc)OH (2.87 g, 11.6 mmol), $NaHCO_3$ (2.44 g, 29.1 mmol) in 1:1 THF/H_2O (40 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.16 (br s, 1H, CO_2H), 8.20-8.14 (m, 2H, 2 x N*H*), 8.06 (s, 1H, C*H*O), 6.76 (t, *J* = 4.9 Hz, 1H, εN*H*-Lys), 4.19-4.12 (m, 1H, *αH*-Lys), 3.79 (d, *J* = 7.5 Hz, 2H, *αH*-Gly), 2.91-2.85 (m, 2H, εCH_2-Lys), 1.71-1.52 (m, 2H, βCH_2-Lys), 1.39-1.33 (m, 11H, ^{t}Bu and δCH_2-Lys), 1.31-1.22 (m, 2H, γCH_2-Lys); HRMS for $C_{14}H_{25}N_3O_6$ $[M+Na]^+$ calcd.: 354.1636, found: 354.1635.

***N*-For-L-Met-L-Lys(Boc)OH (2.3e):** Following Method (a), dipeptide **2.3e** was obtained as a white solid (1.25 g, 84%), in two steps, starting from *N*-For-L-Met-OH (0.65 g, 3.67 mmol), *N*-

hydroxysuccinimide (0.51 g, 4.40 mmol), DCC (0.98 g, 4.77 mmol) in THF (20 mL) and H-L-Lys(Boc)-OH (0.99 g, 4.03 mmol), $NaHCO_3$ (1.23 g, 14.7 mmol) in 1:1 THF/H_2O (40 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.56 (br s, 1H, CO_2H), 8.28 (d, J = 8.3 Hz, 1H, N*H*-Met), 8.24 (d, J = 7.6 Hz, 1H, αN*H*-Lys), 8.00 (s, 1H, C*H*O), 6.76 (t, J = 5.3 Hz, 1H, εN*H*-Lys), 4.50-4.43 (m, 1H, *αH*-Met), 4.15-4.08 (m, 1H, *αH*-Lys), 2.91-2.85 (m, 2H, εCH_2-Lys), 2.44 (t, J = 8.0 Hz, 2H, γCH_2-Met), 2.03 (s, 3H, -SCH_3), 1.95-1.77 (m, 2H, βCH_2-Met), 1.75-1.56 (m, 2H, βCH_2-Lys), 1.36-1.21 (m, 13H, ^{t}Bu, δCH_2-Lys and γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 173.4 (CO-Lys), 170.7 (CO-Met), 160.8 (CHO), 155.5 (OCONH), 77.2 (1C of ^{t}Bu), 51.9 (αC-Lys), 50.1 (αC-Met), 39.6 (εC-Lys), 32.4 (βC-Met), 30.4 (βC-Lys), 29.2 (γC-Met), 29.1 (δC-Lys), 28.2 (CH_3-^{t}Bu), 22.8 (γC-Lys), 14.6 (SCH_3); HRMS for $C_{17}H_{31}N_3O_6S$ $[M-H]^-$ calcd.: 404.1861, found: 404.1861.

***N*-For-L-Met-L-Lys(Boc)OMe (2.3f):** Following Method (b), dipeptide **2.3f** was obtained as a white solid (1.02 g, 86%), in two steps, starting from *N*-For-L-Met-OH (0.5 g, 2.82 mmol), *N*-hydroxysuccinimide (0.39 g, 3.39 mmol), DCC (0.76 g, 3.67 mmol) in THF (18 mL) and H-L-Lys(Boc)OMe·HCl salt (0.92 g, 3.10 mmol), $NaHCO_3$ (0.83 g, 9.87 mmol) in 1:1 THF/H_2O (30 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.39 (d, J = 7.2 Hz, 1H, αN*H*-Lys), 8.29 (d, J = 8.2 Hz, 1H, N*H*-Met), 8.01 (s, 1H, C*H*O), 6.76 (t, J = 5.4 Hz, 1H, εN*H*-Lys), 4.51-4.44 (m, 1H, *αH*-Met), 4.22-4.16 (m, 1H, *αH*-Lys), 3.62 (s, 3H, OCH_3), 2.91-2.85 (m, 2H, εCH_2-Lys), 2.44 (t, J = 7.9 Hz, 2H, γCH_2-Met), 2.04 (s, 3H, SCH_3), 1.93-1.77 (m, 2H, βCH_2-Met), 1.75-1.58 (m, 2H, βCH_2-Lys), 1.37 (s, 9H, ^{t}Bu), 1.35-1.21 (m, 4H, δCH_2-Lys and γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.4 (CO-Lys), 170.9 (CO-Met), 160.8 (CHO), 155.5 (OCONH), 77.3 (1C of ^{t}Bu), 52.0 (αC-Lys), 51.8 (OCH_3), 50.0 (αC-Met), 39.6 (εC-Lys), 32.3 (βC-Met), 30.3 (βC-Lys), 29.2 (γC-Met), 29.0 (δC-Lys), 28.2 (CH_3-^{t}Bu), 22.7 (γC-Lys), 14.6 (SCH_3); HRMS for $C_{18}H_{33}N_3O_6S$ $[M-H]^-$ calcd.: 418.2017, found: 418.2022.

Boc-L-Ala-L-Lys(Cbz)OH (2.11): Following Method (a), dipeptide **2.11** was obtained as a white solid (1.63 g, 68%), in two steps, starting from Boc-L-Ala-OH (1.0 g, 5.29 mmol), *N*-hydroxysuccinimide (0.73 g, 6.35 mmol), DCC (1.42 g, 6.88 mmol) in THF (25 mL) and H-L-Lys(Cbz)OH (1.70 g, 6.08 mmol), $NaHCO_3$ (1.78 g, 21.2 mmol) in 1:1 THF/H_2O (50 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 10.49 (br s, 1H, CO_2H), 7.80 (d, J = 7.6 Hz, 1H, αN*H*-Lys), 7.29-7.23 (m, 5H, Ar-*H*), 7.10 (t, J = 5.9 Hz, 1H, εN*H*-Lys), 6.77 (d, J = 7.5 Hz, 1H, N*H*-Ala), 4.91 (s, 2H, OCH_2Ph), 4.10-4.03 (m, 1H, *αH*-Lys), 3.93-3.86 (m, 1H, *αH*-Ala), 2.90-2.85 (m, 2H, εCH_2-Lys), 1.64-1.41 (m, 2H, βCH_2-Lys), 1.35-1.20 (m, 11H, CH_3-^{t}Bu and δCH_2-Lys), 1.16-0.93 (m, 5H, CH_3-Ala and γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 173.6 (CO-Lys), 172.9 (CO-Ala), 156.2 (OCONH), 155.1 (OCONH), 137.4 (1C-Ph), 128.5 (Ar-C), 127.8 (Ar-C), 78.2 (1C of ^{t}Bu), 65.2 (-OCH_2Ph), 51.7 (αC-Lys), 49.6 (αC-Ala), 40.2 (εC-Lys), 30.9 (βC-Lys), 29.1 (δC-Lys), 28.3 (CH_3-^{t}Bu), 22.6 (γC-Lys), 18.2 (CH_3-Ala); HRMS for $C_{22}H_{33}N_3O_7$ $[M-H]^-$ calcd.: 450.2245, found: 450.2236.

***N*-For-Met-L-Glu-OMe (2.2a):** To a stirring solution of **2.1a** (1.00 g, 2.44 mmol) in EtOH was added 10% Pd/C Degussa (0.5 g, 50% w/w) and the mixture was hydrogenated at atmospheric pressure using a balloon filled with H_2 for 16 h. The catalyst was removed by filtration and the filtrate was concentrated under reduced pressure. The resulting crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 94:6, v/v) to give compound **2.2a** (0.75 g, 96%) as a white solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.15 (br s, 1H, CO_2H), 8.43 (t, J = 7.4 Hz, 1H, N*H*-Glu), 8.30 (d, J = 7.9 Hz, 1H, N*H*-Met), 8.01 (s, 1H, C*H*O), 4.89-4.41 (m, 1H, *αH*-Met), 4.31-4.23 (m, 1H, *αH*-Glu), 3.62 (s, 3H, OCH_3), 2.44 (t, J = 7.9 Hz, 2H, γCH_2-Met), 2.29 (t, J = 7.5 Hz, 2H, γCH_2-Glu), 2.04 (s, 3H, -SCH_3), 2.01-1.72 (unresolved m, 4H, βCH_2-

Glu and βC*H*$_2$-Met); ^{13}C NMR (75 MHz, DMSO-*d*$_6$) δ = 173.6 (δCO-Glu), 172.0 (*α*CO-Glu), 171.0 (CO-Met), 160.9 (CHO), 51.9 (OCH$_3$), 51.2 (αC-Glu), 50.1 (αC-Met), 32.1 (βC-Met), 29.8 (γC-Glu), 29.2 (γC-Met), 25.9 (βC-Glu), 14.6 (SCH$_3$); HRMS for $C_{11}H_{20}N_2O_6S$ $[M-H]^-$ calcd.: 319.0969, found: 319.0975.

***N*-For-Met-L-Glu-NH$_2$ (2.2b):** Following a similar procedure as used for the synthesis of compound **2.2a**, **2.1b** (0.76 g, 2.44 mmol) was hydrogenated in the presence of 10% Pd/C Degussa (0.38 g, 50% w/w) in EtOH and the resulting product was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:2, v/v; 97:4, v/v; 94:8, v/v) to give **2.2b** (0.51 g, 86%) as a white solid. ^{1}H NMR (300 MHz, DMSO-*d*$_6$) *δ* = 12.09 (br s, 1H, CO$_2$*H*), 8.33 (d, *J* = 7.9 Hz, 1H, N*H*-Met), 8.04-8.01 (m, N*H*-Glu and C*H*O), 7.29 (br s, 1H, N*H*-CONH$_2$), 7.07 (br s, 1H, N*H*-CONH$_2$), 4.44-4.37 (m, 1H, *αH*-Met), 4.21-4.14 (m, 1H, *αH*-Glu), 2.44 (t, *J* = 7.9 Hz, 2H, γC*H*$_2$-Met), 2.29 (t, *J* = 7.6 Hz, 2H, γC*H*$_2$-Glu), 2.03 (s, 3H, SC*H*$_3$), 1.97-1.85 (m, 2H, βC*H*$_2$-Glu), 1.81-1.71 (m, 2H, βC*H*$_2$-Met); ^{13}C NMR (75 MHz, DMSO-*d*$_6$) δ = 174.0 (δCO-Glu), 172.9 (αCO-Glu), 170.6 (CO-Met), 161.2 (CHO), 51.8 (αC-Glu), 50.6 (αC-Met), 31.9 (βC-Met), 30.3 (γC-Glu), 29.3 (γC-Met), 27.2 (βC-Glu), 14.6 (SCH$_3$); HRMS for $C_{11}H_{19}N_3O_5S$ $[M-H]^-$ calcd.: 304.0972, found: 304.0975.

***N*-For-L-Met-L-Lys(NH$_2$·TFA)-OMe (2.8):** To a stirring solution of **2.3f** 162.9 mg, 0.388 mmol), thioanisole (0.046 mL, 0.388 mmol) in CH_2Cl_2 (4 mL) was slowly added trifluoroacetic acid (1 mL) at 0 °C. The mixture was then slowly warmed to r.t. and stirred for 4 h. After removal of all the volatiles under reduced pressure, the resulting residue was co-evaporated with toluene (3 times) to afford compound **2.8** as a colorless foam (168.3 mg, quantitative), which was used directly in the next step without any further purification.

***N*-For-Gly-L-Glu(OBn)-L-Ala-OMe (2.4a):** Following Method (b), tripeptide **2.4a** was obtained as a white solid (2.04 g, 79%), in two steps, starting from **2.3a** (2.04 g, 6.33 mmol), *N*-hydroxysuccinimide (0.87 g, 7.59 mmol), DCC (1.70 g, 8.23 mmol) in THF (36 mL) and H-L-Ala-OMe·HCl salt (1.02 g, 7.28 mmol), $NaHCO_3$ (2.13 g, 25.32 mmol) in 1:1 THF/H_2O (40 mL). ^{1}H NMR (300 MHz, DMSO-*d*$_6$) *δ* = 8.43 (d, *J* = 6.8 Hz, 1H, N*H*-Ala), 8.21 (t, *J* = 5.2 Hz, 1H, N*H*-Gly), 8.10 (d, *J* = 8.1 Hz, 1H, N*H*-Glu), 8.06 (s, 1H, C*H*O), 7.38-7.32 (m, 5H, Ar-*H*), 5.09 (s, 2H, OC*H*$_2$Ph), 4.40-4.33 (m, 1H, αC*H*-Glu), 4.30-4.21 (m, 1H, αC*H*-Ala), 3.77 (d, *J* = 7.3 Hz, 2H, αC*H*$_2$-Gly), 3.59 (s, 3H, OC*H*$_3$), 2.41 (t, *J* = 8.1 Hz, 2H, γC*H*$_2$-Glu), 2.02-1.75 (m, 2H, βC*H*$_2$-Glu), 1.29 (d, *J* = 7.3 Hz, 3H, C*H*$_3$-Ala); ^{13}C NMR (75 MHz, DMSO-d$_6$) δ = 172.8 (δCO-Glu), 172.2 (CO-Ala), 170.8 (αCO-Glu), 168.3 (CO-Gly), 161.4 (CHO), 136.2 (1C Ph), 128.4 (Ar-C), 128.0 (Ar-C), 127.9 (Ar-C), 65.5 (OCH$_2$Ph), 51.8 (OCH$_3$), 51.2 (αC-Glu), 47.6 (αC-Ala), 40.5 (αC-Gly), 29.8 (γC-Glu), 27.5 (βC-Glu), 16.6 (CH$_3$-Ala); HRMS for $C_{19}H_{25}N_3O_7$ $[M+H]^+$ calcd.: 408.1765, found: 408.1765.

***N*-For-Gly-L-Glu(O^{t}Bu)-L-Phe-OMe (2.4b):** Following Method (a), tripeptide **2.4b** was obtained as a white solid (2.29 g, 69%), in two steps, starting from **2.3b** (2.11 g, 7.32 mmol), *N*-hydroxysuccinimide (1.01 g, 8.78 mmol), DCC (1.96 g, 9.51 mmol) in THF (14 mL) and H-L-Phe-OMe·HCl salt (2.00 g, 9.27 mmol), $NaHCO_3$ (1.72 g, 20.5 mmol, 2.2 eq.) in 1:1 THF/H_2O (40 mL). ^{1}H NMR (300 MHz, MeOD) *δ* = 8.18 (s, 1H, C*H*O), 7.29-7.20 (m, 5H, Ar-*H*), 4.67 (dd, *J* = 8.5, 5.6 Hz, 1H, αC*H*-Phe), 4.40 (dd, *J* = 8.3, 5.3 Hz, 1H, αC*H*-Glu), 3.92 (br s, 2H, αC*H*$_2$-Gly), 3.70 (s, 3H, OC*H*$_3$), 3.21-3.15 (m, 1H, βC*H*$_2$-Phe), 3.06-2.98 (m, 1H, βC*H*$_2$-Phe), 2.31-2.26 (m, 2H, γC*H*$_2$-Glu), 2.04-2.00 (m, 1H, βC*H*$_2$-Glu), 1.88-1.85 (m, 1H, βC*H*$_2$-Glu), 1.46 (s, 9H, C(C*H*$_3$)$_3$); ^{13}C NMR (75 MHz, MeOD) δ = 172.2 (δCO-Glu), 171.6 (CO), 171.4 (CO), 169.1 (CO), 162.5 (CHO), 136.3 (1C-Ph), 128.5 (Ar-C), 127.8 (Ar-C), 126.2 (Ar-C), 53.5, 52.0, 51.0, 40.3, 36.5,

30.7 (γC-Glu), 26.6 (βC-Glu), 26.5 (^{t}Bu); HRMS for $C_{22}H_{32}N_3O_7$ $[M+H]^+$ calcd.: 450.2234, found: 450.2233.

***N*-For-Gly-L-β-amino-Ala(Fmoc)-L-Phe-OMe (2.4c):** Following Method (b), tripeptide **2.4c** was obtained as a white solid (1.42 g, 72.8%), in two steps, starting from **2.3c** (1.4 g, 3.40 mmol), *N*-hydroxysuccinimide (0.47 g, 4.09 mmol), DCC (0.91 g, 4.42 mmol) in THF (25 mL) and H-L-Phe-OMe·HCl salt (0.84 g, 3.91 mmol), $NaHCO_3$ (1.14 g, 13.6 mmol) in 1:1 THF/H_2O (50 mL). ^{1}H NMR (500 MHz, DMSO-d_6) δ = 8.41 (d, J = 7.6 Hz, 1H, N*H*-Phe), 8.24 (t, J = 4.9 Hz, 1H, N*H*-Gly), 8.09 (s, 1H, CHO), 8.02 (d, J = 8.1 Hz, 1H, N*H*-β-amino-Ala), 7.88 (d, J = 7.6 Hz, 2H, Ar*H*-Fmoc), 7.69 (d, J = 7.5 Hz, 2H, Ar*H*-Fmoc), 7.43-7.15 (m, 9H, Ar*H*-Fmoc and Ar*H*-Phe), 7.17 (t, J = 5.0 Hz, 1H, βN*H*-β-amino-Ala), 4.51-4.47 (m, 1H, *αH*-Phe), 4.55-4.42 (m, 1H, *αH*-β-amino-Ala), 4.30-4.28 (m, 1H, OCH_2-Fmoc), 4.24-4.21 (m, 1H, C*H*-Fmoc), 3.79 (d, J = 5.7 Hz, 2H, αCH_2-Gly), 3.56 (s, 3H, OCH_3), 3.35-3.15 (m, 2H, βCH_2-β-amino-Ala), 3.06-2.93 (m, 2H, βCH_2-Phe); ^{13}C NMR (125 MHz, DMSO-d_6) δ = 171.7 (CO-Phe), 169.7 (CO-β-amino-Ala), 168.5 (CO-Gly), 161.6 (CHO), 156.3 (OCONH), 143.9 (1C Fmoc), 140.8 (1C Fmoc), 137.1 (1C Ph), 129.2 (Ar-C), 129.1 (Ar-C), 128.3 (Ar-C), 127.7 (Ar-C), 127.4 (Ar-C), 127.2 (Ar-C), 126.7 (Ar-C), 125.3 (Ar-C), 121.5 (Ar-C), 120.1 (Ar-C), 65.7 (OCH_2-Fmoc), 53.8 (αC-Phe), 52.7 (αC-β-amino-Ala), 52.0 (OCH_3), 46.7 (CH-Fmoc), 42.3 (βC-Phe), 40.7 (αC-Gly), 36.6 (βC-β-amino-Ala); HRMS for $C_{31}H_{32}N_4O_7$ $[M+H]^+$ calcd.: 573.2343, found: 573.2346.

***N*-For-Gly-L-Lys(Boc)-L-Phe-OMe (2.4d):** Following Method (b), tripeptide **2.4d** was obtained as a white solid (1.82 g, 72%), in two steps, starting from **2.3d** (1.7 g, 5.13 mmol), *N*-hydroxysuccinimide (0.71 g, 6.16 mmol), DCC (1.38 g, 6.67 mmol) in THF (35 mL) and H-L-Phe-OMe·HCl salt (1.33 g, 6.16 mmol), $NaHCO_3$ (1.72 g, 20.52 mmol) in 1:1 THF/H_2O (50 mL). ^{1}H NMR (500 MHz, DMSO-d_6) δ = 8.37 (d, J = 7.4 Hz, 1H, N*H*-Phe), 8.19 (t, J = 5.2 Hz, 1H, N*H*-Gly), 8.06 (s, 1H, C*H*O), 7.97 (d, J = 8.1 Hz, 1H, N*H*-Lys), 7.30-7.19 (m, 5H, Ar-*H*), 6.73 (t, J = 5.1 Hz, 1H, εN*H*-Lys), 4.49-4.41 (m, 1H, *αH*-Phe), 4.31-4.24 (m, 1H, *αH*-Lys), 3.76 (d, J = 5.7 Hz, 2H, *αH*-Gly), 3.57 (s, 3H, OCH_3), 3.06-2.92 (m, 2H, βCH_2-Phe), 2.89-2.83 (m, 2H, εCH_2-Lys), 1.61-1.43 (m, 2H, βCH_2-Lys), 1.37-1.29 (m, 11H, ^{t}Bu and δCH_2-Lys), 1.26-1.13 (m, 2H, γCH_2-Lys); ^{13}C NMR (125 MHz, DMSO-d_6) δ = 171.7 (CO-Phe), 171.6 (CO-Lys), 168.0 (CO-Gly), 161.4 (CHO), 155.5 (OCONH), 137.1 (1C-Ph), 129.0 (Ar-C), 128.2 (Ar-C), 126.5 (Ar-C), 77.3 (1C of ^{t}Bu), 53.6 (αC-Phe), 52.1 (αC-Lys), 51.8 (OCH_3), 40.4 (αC-Gly), 39.6 (εC-Lys), 36.5 (βC-Phe), 31.9 (βC-Lys), 29.3 (δC-Lys), 28.3 (CH_3-^{t}Bu), 22.4 (γC-Lys); HRMS for $C_{24}H_{36}N_4O_7$ $[M+H]^+$ calcd.: 493.2656, found: 493.2659.

***N*-For-L-Met-L-Lys(Boc)-L-Ala-OMe (2.4e):** Following Method (b), tripeptide **2.4e** was obtained as a white solid (1.50 g, 88.5%), in two steps, starting from **2.3e** (1.4 g, 3.45 mmol), *N*-hydroxysuccinimide (0.48 g, 4.14 mmol), DCC (0.93 g, 4.49 mmol) in THF (30 mL) and H-L-Ala-OMe·HCl salt (0.53 g, 3.80 mmol), $NaHCO_3$ (0.87 g, 10.36 mmol) in 1:1 THF/H_2O (40 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.30 (t, J = 6.9 Hz, 1H, N*H*), 8.27 (d, J = 9.4 Hz, 1H, N*H*), 8.02-8.00 (m, 2H, C*H*O and N*H*), 6.73 (t, J = 5.3 Hz, 1H, εN*H*-Lys), 4.47-4.40 (m, 1H, *αH*-Met), 4.29-4.20 (m, 2H, *αH*-Lys and *αH*-Ala), 3.61 (s, 3H, OCH_3), 2.91-2.84 (m, 2H, εCH_2-Lys), 2.42 (t, J = 7.9 Hz, 2H, γCH_2-Met), 2.02 (s, 3H, SCH_3), 1.94-1.72 (m, 2H, βCH_2-Met), 1.69-1.49 (m, 2H, βCH_2-Lys), 1.40-1.31 (m, 11H, CH_3-^{t}Bu and δCH_2-Lys), 1.29-1.20 (m, 5H, CH_3-Ala and γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.9 (CO-Ala), 171.4 (CO-Lys), 170.5 (CO-Met), 160.9 (CHO), 155.5 (OCONH), 77.3 (1C of ^{t}Bu), 52.2 (αC-Lys), 51.8 (OCH_3), 50.3 (αC-Met), 47.5 (αC-Ala), 39.8 (εC-Lys), 32.2 (βC-Met), 31.7 (βC-Lys), 29.3 (γC-Met), 29.2 (δC-Lys), 28.2 (CH_3-^{t}Bu), 22.5 (γC-Lys), 16.8 (CH_3-Ala), 14.6 (SCH_3); HRMS for $C_{21}H_{38}N_4O_7S$ $[M+Na]^+$ calcd.: 513.2353, found: 513.2352.

***N*-For-L-Met-L-Lys(Boc)-L-Lys(Cbz)-OMe (2.4f):** Following Method (b), tripeptide **2.4f** was obtained as a white solid (1.07 g, 67%), in two steps, starting from **3e** (0.92 g, 2.27 mmol), *N*-hydroxysuccinimide (0.31 g, 2.72 mmol), DCC (0.61 g, 2.95 mmol) in THF (20 mL) and H-L-Lys(Cbz)-OMe·HCl salt (0.86 g, 2.61 mmol), $NaHCO_3$ (0.76 g, 9.08 mmol) in 1:1 THF/H_2O (30 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.31-8.25 (m, 2H, 2 x N*H*), 8.02-8.00 (m, 2H, C*H*O and N*H*), 7.36-7.30 (m, 5H, Ar-*H*), 7.22 (t, *J* = 5.1 Hz, 1H, εN*H*-Lys$_{C\text{-}ter}$), 6.73 (t, *J* = 5.0 Hz, 1H, εN*H*-Lys$_{N\text{-}ter}$), 5.00 (s, 2H, OCH_2Ph), 4.47-4.40 (m, 1H, *αH*-Met), 4.32-4.25 (m, 1H, *αH*-Lys$_{N\text{-}ter}$), 4.22-4.15 (m, 1H, *αH*-Lys$_{C\text{-}ter}$), 3.61 (s, 3H, OCH_3), 2.99-2.93 (m, 2H, εCH_2-Lys), 2.89-2.83 (m, 2H, εCH_2-Lys), 2.41 (t, *J* = 7.9 Hz, 2H, γCH_2-Met), 2.02 (s, 3H, SCH_3), 1.94-1.70 (m, 2H, βCH_2-Met), 1.69-1.44 (m, 4H, 2 x βCH_2-Lys), 1.40-1.17 (m, 17H, CH_3-^{t}Bu, 2 x δCH_2-Lys and 2 x γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.4 (CO-Lys$_{C\text{-}ter}$), 171.6 (CO-Met), 170.4 (CO-Lys$_{N\text{-}ter}$), 161.0 (CHO), 156.0 (OCONH), 155.5 (OCONH), 137.3 (1C-Ph), 128.3 (Ar-C), 127.7 (Ar-C), 77.3 (1C of ^{t}Bu), 65.1 (OCH_2Ph), 52.3 (αC-Lys$_{N\text{-}ter}$), 51.7 (OCH_3 and αC-Lys$_{C\text{-}ter}$), 50.4 (αC-Met), 40.0 (εC-Lys), 32.1 (βC-Met), 31.9 (βC-Lys), 30.5 (βC-Lys), 29.3 (γC-Met), 29.2 (δC-Lys), 28.8 (δC-Lys), 28.2 (CH_3-^{t}Bu), 22.6 (γC-Lys), 22.5 (γC-Lys), 14.5 (SCH_3); HRMS for $C_{32}H_{51}N_5O_9S$ $[M-H]^-$ calcd.: 680.3334, found: 680.3336.

***N*-For-L-Met-L-Lys(Boc)-L-Phe-OMe (2.4g):** Following Method (b), tripeptide **2.4g** was obtained as a white solid (1.02 g, 86%), in two steps, starting from **2.3e** (0.5 g, 1.23 mmol), *N*-hydroxysuccinimide (0.17 g, 1.48 mmol), DCC (0.33 g, 1.60 mmol) in THF (18 mL) and H-L-Phe-OMe·HCl salt (0.29 g, 1.36 mmol), $NaHCO_3$ (0.41 g, 4.93 mmol) in 1:1 THF/H_2O (30 mL). ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.14 (s, 1H, C*H*O), 7.69 (d, *J* = 8.1 Hz, 1H, αN*H*-Lys), 7.60-7.57 (m, 2H, N*H*-Met and N*H*-Phe), 7.27-7.11 (m, 5H, Ar-*H* Phe), 5.03-4.98 (m, 1H, εN*H*-Lys), 4.96-4.88 (m, 1H, *αH*-Met), 4.84-4.77 (m, 1H, *αH*-Phe), 4.74-4.69 (m, 1H, *αH*-Lys), 3.69 (s, 3H, OCH_3), 3.11-3.01 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.47 (t, *J* = 6.8 Hz, 2H, γCH_2-Met), 2.04 (s, 3H, SCH_3), 2.00-1.90 (m, 2H, βCH_2-Met), 1.79-1.63 (m, 2H, βCH_2-Lys), 1.47-1.40 (m, 11H, ^{t}Bu and δCH_2-Lys), 1.35-1.28 (m, 2H, γCH_2-Lys); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.0 (CO-Phe), 171.5 (CO-Lys), 171.1 (CO-Met), 161.4 (CHO), 156.2 (OCONH), 136.0 (1C of Phe), 129.2 (Ar-C), 128.6 (Ar-C), 127.2 (Ar-C), 79.1 (1C of ^{t}Bu), 53.7 (αC-Phe), 52.9 (αC-Lys), 52.4 (OCH_3), 50.6 (αC-Met), 40.2 (εC-Lys), 37.9 (βC-Phe), 32.8 (βC-Met and βC-Lys), 30.0 (γC-Met), 29.6 (δC-Lys), 28.5 (CH_3-^{t}Bu), 22.6 (γC-Lys), 15.4 (SCH_3); HRMS for $C_{27}H_{42}N_4O_7S$ $[M-H]^-$ calcd.: 565.2701, found: 565.2700.

***N*-For-L-Met-L-Lys(Boc)-L-Lys(Boc)-OH (2.4h):** Following Method (a), tripeptide **2.4h** was obtained as a white solid (0.54 g, 62%), in two steps, starting from **2.3e** (0.56 g, 1.38 mmol), *N*-hydroxysuccinimide (0.19 g, 1.66 mmol), DCC (0.37 g, 1.79 mmol) in THF (15 mL) and H-L-Lys(Boc)-OH (0.39 g, 1.59 mmol), $NaHCO_3$ (0.46 g, 5.52 mmol) in 1:1 THF/H_2O (20 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.51 (s, 1H, CO_2H), 8.28 (d, *J* = 8.0 Hz, 1H, N*H*-Met), 8.04-8.02 (m, 3H, 2 x N*H*-Lys and C*H*O), 6.77-6.71 (m, 2H, 2 x εN*H*-Lys), 4.47-4.40 (m, 1H, *αH*-Met), 4.28-4.24 (m, 1H, *αH*-Lys$_{N\text{-}ter}$), 4.23-4.21 (m, 1H, *αH*-Lys$_{C\text{-}ter}$), 2.91-2.83 (m, 4H, 2 x εCH_2-Lys), 2.42 (t, *J* = 7.8 Hz, 2H, γCH_2-Met), 2.02 (s, 3H, SCH_3), 1.90-1.70 (m, 2H, βCH_2-Met), 1.69-1.48 (m, 4H, 2 x βCH_2-Lys), 1.40-1.15 (m, 26H, 2 x CH_3-^{t}Bu, 2 x δCH_2-Lys and 2 x γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 173.4 (CO-Lys$_{C\text{-}ter}$), 171.5 (CO-Met), 170.5 (CO-Lys$_{N\text{-}ter}$), 160.9 (CHO), 155.5 (2 x OCONH), 77.3 (1C of ^{t}Bu), 52.3 (αC-Lys$_{N\text{-}ter}$), 51.8 (αC-Lys$_{C\text{-}ter}$), 50.3 (αC-Met), 39.6 (2 x εC-Lys), 32.2 (βC-Met), 31.6 (βC-Lys), 30.6 (βC-Lys), 29.3 (γC-Met), 29.1 (2 x δC-Lys), 28.2 (CH_3-^{t}Bu), 22.6 (2 x γC-Lys), 14.6 (SCH_3); HRMS for $C_{28}H_{51}N_5O_9S$ $[M-H]^-$ calcd.: 632.3334, found: 632.3334.

***N*-For-L-Met-L-Lys(Boc)-L-Ala-OH (2.4i):** Following Method (a), tripeptide **2.4i** was obtained as a white solid (2.04 g, 85%), in two steps, starting from **2.3e** (2.04 g, 5.03 mmol), *N*-

hydroxysuccinimide (0.70 g, 6.04 mmol), DCC (1.35 g, 6.54 mmol) in THF (40 mL) and H-L-Ala-OH (0.49 g, 5.53 mmol), $NaHCO_3$ (1.69 g, 20.12 mmol) in 1:1 THF/H_2O (50 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.52 (br s, 1H, CO_2H), 8.28 (d, J = 8.0 Hz, 1H, αN*H*-Met), 8.13 (d, J = 7.1 Hz, 1H, αN*H*-Ala), 8.03-8.01 (m, 2H, C*H*O and αN*H*-Lys), 6.73 (t, J = 4.4 Hz, 1H, εN*H*-Lys), 4.47-4.40 (m, 1H, *αH*-Met), 4.26-4.15 (m, 2H, *αH*-Ala and *αH*-Lys), 2.89-2.84 (m, 2H, εCH_2-Lys), 2.43 (t, J = 7.8 Hz, 2H, γCH_2-Met), 2.03 (s, 3H, -SCH_3), 1.93-1.70 (m, 2H, βCH_2-Met), 1.63-1.48 (m, 2H, βCH_2-Lys), 1.40-1.33 (m, 11H, ^{t}Bu and δCH_2-Lys), 1.30-1.19 (m, 5H, CH_3-Ala and γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.0 (CO_2H), 171.2 (CO-Lys), 170.5 (CO-Met), 161.0 (CHO), 155.5 (OCONH), 77.3 (1C of ^{t}Bu), 52.2 (αC-Lys), 50.4 (αC-Met), 47.4 (αC-Ala), 40.2 (εC-Lys), 32.2 (βC-Met), 31.7 (βC-Lys), 29.3 (γC-Met), 29.2 (δC-Lys), 28.3 (CH_3-^{t}Bu), 22.6 (γC-Lys), 17.1 (CH_3-Ala), 14.6 (SCH_3); HRMS for $C_{20}H_{36}N_4O_7S$ $[M-H]^-$ calcd.: 475.2232, found: 475.2231.

Boc-L-Ala-L-Lys(Cbz)-L-Ala-OMe (2.12): Following Method (b), tripeptide **2.12** was obtained as a white solid (1.10 g, 75%), in two steps, starting from **2.11** (1.23 g, 2.72 mmol), *N*-hydroxysuccinimide (0.38 g, 3.27 mmol), DCC (0.73 g, 3.54 mmol) in THF (20 mL) and H-L-Ala-OMe·HCl (0.44 g, 3.13 mmol), $NaHCO_3$ (0.92 g, 10.92 mmol) in 1:1 THF/H_2O (30 mL). ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.33-7.29 (m, 5H, Ar-*H*), 6.99-6.94 (m, 2H, 2 x N*H*), 5.28-5.22 (m, 2H, 2 x N*H*), 5.08 (s, 2H, OCH_2Ph), 4.57-4.44 (m, 2H, *αH*-Lys and *αH*-$Ala_{C\text{-}ter}$), 4.20-4.16 (m, 1H, *αH*-$Ala_{N\text{-}ter}$), 3.69 (s, 3H, OCH_3), 3.21-3.14 (m, 2H, εCH_2-Lys), 1.89-1.61 (m, 2H, βCH_2-Lys), 1.55-1.47 (m, 2H, δCH_2-Lys), 1.42 (s, 9H, CH_3-^{t}Bu), 1.39-1.36 (m, 5H, CH_3-$Ala_{C\text{-}ter}$ and γCH_2-Lys), 1.32 (d, J = 7.0 Hz, 3H, CH_3-$Ala_{N\text{-}ter}$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 173.3 (CO-$Ala_{C\text{-}ter}$), 172.9 (CO-$Ala_{N\text{-}ter}$), 173.3 (CO-Lys), 156.8 (OCONH), 155.7 (OCONH), 136.7 (1C-Ph), 128.6 (Ar-C), 128.3 (Ar-C), 128.2 (Ar-C), 80.3 (1C of ^{t}Bu), 66.7 (OCH_2Ph), 52.8 (αC-Lys), 52.5 (OCH_3), 50.3 (αC-$Ala_{N\text{-}ter}$), 48.2 (αC-$Ala_{C\text{-}ter}$), 40.4 (εC-Lys), 31.9 (βC-Lys), 29.3 (δC-Lys), 28.4 (CH_3-^{t}Bu), 22.2 (γC-Lys), 18.4 (CH_3-$Ala_{N\text{-}ter}$), 18.0 (CH_3-$Ala_{C\text{-}ter}$); HRMS for $C_{26}H_{40}N_4O_8$ $[M+H]^+$ calcd.: 537.2919, found: 537.2916.

***N*-For-Gly-L-Glu-L-Ala-OMe (2.5a):** Following a similar procedure as used for the synthesis of **2.2a**, compound **2.4a** (2.00 g, 4.91 mmol) was hydrogenated in the presence of 10% Pd/C Degussa (0.20 g, 10% w/w) in MeOH (60 mL) to give **2.5a** (1.42 g, 91%) as a white solid after purification. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.09 (br s, 1H, γCO_2H-Glu), 8.41 (d, J = 6.8 Hz, 1H, N*H*-Ala), 8.21 (t, J = 5.2 Hz, 1H, N*H*-Gly), 8.09-8.06 (m, 2H, N*H*-Glu and C*H*O), 4.36-4.29 (m, 1H, αC*H*-Glu), 4.27-4.20 (m, 1H, αC*H*-Ala), 3.77 (d, J = 5.8 Hz, 2H, αCH_2-Gly), 3.61 (s, 3H, OCH_3), 2.25 (t, J = 8.1 Hz, 2H, γCH_2-Glu), 1.95-1.67 (m, 2H, βCH_2-Glu), 1.28 (d, J = 7.0 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 174.0 (δCO-Glu), 172.9 (CO-Ala), 171.0 (αCO-Glu), 168.3 (CO-Gly), 161.5 (CHO), 51.9 (OCH_3), 51.4 (αC-Glu), 47.6 (αC-Ala), 40.5 (αC-Gly), 29.9 (γC-Glu), 27.6 (βC-Glu), 16.7 (CH_3-Ala); HRMS for $C_{12}H_{19}N_3O_7$ $[M-H]^-$ calcd.: 316.1150, found: 316.1152.

***N*-For-Gly-L-Glu-L-Phe-OMe (2.5b):** A solution of **2.4b** (0.13 g, 0.29 mmol) in CH_2Cl_2 (4 mL) was cooled to 0 °C and trifluoroacetic acid was then added. The reaction mixture was then stirred for 15 min at 0 °C and subsequently for 2 h at r.t. To this solution was then added toluene (6 mL), after which the volatiles were removed *in vacuo*. The remaining crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 98:1, v/v; 96:5, v/v; 94:7, v/v, 94:10, v/v) to afford **2.5b** as a white foam (0.11 g, quant.). ^{1}H NMR (300 MHz, MeOD) δ = 8.08 (s, 1H, C*H*O), 7.21-7.10 (m, 5H, Ar-*H*), 4.61-4.56 (m, 1H, αC*H*-Phe) 4.35 (dd, J = 8.3, 5.6 Hz, 1H, αC*H*-Glu), 3.83 (br s, 2H, αCH_2-Gly), 3.59 (s, 3H, OCH_3), 3.11-3.04 (m, 1H, βCH_2-Phe), 2.96-2.88 (m, 1H, βCH_2-Phe), 2.26 (t, J = 8.0 Hz, 2H, γCH_2-Glu), 1.97-1.93 (m, 1H, βCH_2-Glu), 1.83-1.79 (m, 1H, βCH_2-Glu); ^{13}C NMR (75 MHz, MeOD) δ = 173.5 (δCO-Glu), 170.2 (CO), 169.9 (CO), 167.6 (CO), 161.1

(CHO), 134.8 (1C-Ph), 127.0 (Ar-C), 126.3 (Ar-C), 124.6 (Ar-C), 50.6, 49.5, 46.6, 38.8, 34.9, 27.8 (γC-Glu), 25.1 (βC-Glu); HRMS for $C_{18}H_{22}N_3O_7$ $[M-H]^-$ calcd.: 392.1463, found: 392.1462.

***N*-For-Gly-L-β-amino-Ala-L-Phe-OMe (2.5c):** A solution of **2.4c** (0.92 g, 10.92 mmol) in a 2:1 mixture of Et_3N:CH_2Cl_2 (21 mL) was stirred at r.t. for 72 h. After reaction completion, the mixture was concentrated under reduced pressure and the crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 98:2, v/v; 96:4, v/v; 94:6, v/v) to afford **2.5c** as a colorless foam (0.57 g, 68%). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.45-8.43 (m, 1H, N*H*-Phe), 8.20-8.19 (m, 1H, N*H*-Gly), 8.07-7.99 (m, 2H, C*H*O and N*H*-β-amino-Ala), 7.29-7.15 (m, 5H, Ar*H*-Phe), 4.50-4.42 (m, 1H, *αH*-Phe), 4.31-4.29 (m, 1H, *αH*-β-amino-Ala), 3.69 (d, *J* = 5.7 Hz, 2H, αCH_2-Gly), 3.58 (s, 3H, OCH_3), 3.41 (br s, 2H, βNH_2-β-amino-Ala), 3.33-3.08 (m, 2H, βCH_2-β-amino-Ala), 3.03-2.88 (m, 2H, βCH_2-Phe); HRMS for $C_{16}H_{22}N_4O_5$ $[M+H]^+$ calcd.: 351.1663, found: 351.1667.

***N*-For-Gly-L-Lys(NH_2·TFA)-L-Phe-OMe (2.5d):** Following a similar procedure as used for the synthesis of **2.8**, compound **2.4d** (265.6 mg, 0.539 mmol), thioanisole (0.063 mL, 0.539 mmol) and trifluoroacetic acid (3 mL) in CH_2Cl_2 (9 mL) afforded **2.5d** as a colorless foam (272.9 mg, quantitative), which was used directly in the next step without any further purification.

***N*-For-L-Met-L-Lys(NH_2·TFA)-L-Ala-OMe (2.5e):** Following a similar procedure as used for the synthesis of **2.8**, compound **2.4e** (200 mg, 0.407 mmol), thioanisole (0.048 mL, 0.407 mmol) and trifluoroacetic acid (2 mL) in CH_2Cl_2 (6 mL) afforded **2.5e** as a colorless foam (205.5 mg, quantitative), which was used directly in the next step without any further purification.

***N*-For-L-Met-L-Lys(NH_2·TFA)-L-Lys(Cbz)-OMe (2.5f):** Following a similar procedure as used for the synthesis of **2.8**, compound **2.4f** (400 mg, 0.587 mmol), thioanisole (0.07 mL, 0.587 mmol) and trifluoroacetic acid (3 mL) in CH_2Cl_2 (9 mL) afforded **2.5f** as a colorless foam (408.1 mg, quantitative), which was used directly in the next step without any further purification.

***N*-For-L-Met-L-Lys(NH_2·TFA)-L-Phe-OMe (2.5g):** Following a similar procedure as used for the synthesis of **2.8**, compound **2.4g** (220 mg, 0.388 mmol), thioanisole (0.046 mL, 0.388 mmol) and trifluoroacetic acid (1 mL) in CH_2Cl_2 (4 mL) afforded **2.5g** as a colorless foam (225.4 mg, quantitative), which was used directly in the next step without any further purification.

Boc-L-Ala-L-Lys-L-Ala-OMe (2.13): Following a similar procedure as used for the synthesis of **2.2a**, compound **2.12** (1.10 g, 2.05 mmol) was hydrogenated in the presence of 10% Pd/C Degussa (0.11 g, 10% w/w) in EtOH (40 mL) to give **2.13** as a colorless foam (0.82 g, quantitative), which was used directly in the next step without any further purification. ^{1}H NMR (500 MHz, DMSO-d_6) δ = 8.39 (d, *J* = 6.4 Hz, 1H, N*H*-Ala$_{C\text{-}ter}$), 7.73 (d, *J* = 8.0 Hz, 1H, N*H*-Lys), 6.96 (d, *J* = 7.4 Hz, 1H, N*H*-Ala$_{N\text{-}ter}$), 4.44 (br s, 2H, εNH_2-Lys), 4.29-4.204 (m, 2H, *αH*-Lys and *αH*-Ala$_{C\text{-}ter}$), 4.00-3.92 (m, 1H, *αH*-Ala$_{N\text{-}ter}$), 3.60 (s, 3H, OCH_3), 2.57 (t, *J* = 6.9 Hz, 2H, εCH_2-Lys), 1.69-1.45 (m, 2H, βCH_2-Lys), 1.45-1.36 (m, 11H, δCH_2-Lys and CH_3-^{t}Bu), 1.28-1.21 (m, 5H, CH_3-Ala$_{C\text{-}ter}$ and γCH_2-Lys), 1.15 (d, *J* = 7.1 Hz, 3H, CH_3-Ala$_{N\text{-}ter}$); ^{13}C NMR (125 MHz, $CDCl_3$) δ = 172.9 (CO-Ala$_{C\text{-}ter}$), 172.4 (CO-Ala$_{N\text{-}ter}$), 171.4 (CO-Lys), 155.0 (OCONH), 78.1 (1C of ^{t}Bu), 51.8 (αC-Lys and OCH_3), 49.7 (αC-Ala$_{N\text{-}ter}$), 47.5 (αC-Ala$_{C\text{-}ter}$), 40.4 (εC-Lys), 32.1 (βC-Lys), 30.7 (δC-Lys), 28.1 (CH_3-^{t}Bu), 22.1 (γC-Lys), 18.0 (CH_3-Ala$_{N\text{-}ter}$), 16.8 (CH_3-Ala$_{C\text{-}ter}$); HRMS for $C_{18}H_{34}N_4O_6$ $[M+H]^+$ calcd.: 403.2551, found: 403.2543.

***N*-For-L-Met-L-Lys(Boc)-L-Ala-L-Ala-OH (2.6a):** Following Method (a), tetrapeptide **2.6a** was obtained as a white solid (0.99 g, 86%), in one-pot without filtration of DCU, starting from **2.4i** (1.0 g, 2.10 mmol), *N*-hydroxysuccinimide (0.29 g, 2.52 mmol), DCC (0.56 g, 2.73 mmol) in THF (18 mL)

and H-L-Ala-OH (0.21 g, 2.31 mmol), $NaHCO_3$ (0.70 g, 8.39 mmol) in 1:1 THF/H_2O (40 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 12.55 (br s, 1H, CO_2H), 8.28 (d, 1H, J = 8.0 Hz, αN*H*-Met), 8.10-8.03 (m, 2H, αN*H*-Lys and αN*H*-$Ala_{N\text{-}ter}$), 8.02 (s, 1H, C*H*O), 7.94 (d, 1H, J = 7.3 Hz, αN*H*- $Ala_{C\text{-}ter}$), 6.73 (t, J = 5.3 Hz, 1H, εN*H*-Lys), 4.47-4.40 (m, 1H, *αH*-Met), 4.31-4.27 (m, 1H, *αH*- $Ala_{C\text{-}ter}$), 4.24-4.16 (m, 2H, *αH*-Lys and *αH*- $Ala_{N\text{-}ter}$), 2.89-2.84 (m, 2H, εCH_2-Lys), 2.43 (t, J = 7.8 Hz, 2H, γCH_2-Met), 2.03 (s, 3H, SCH_3), 1.95-1.73 (m, 2H, βCH_2-Met), 1.68-1.44 (m, 2H, βCH_2-Lys), 1.40-1.33 (m, 11H, ^{t}Bu and δCH_2-Lys), 1.32-1.17 (m, 8H, 2 x CH_3-Ala and γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 173.9 (CO-$Ala_{C\text{-}ter}$), 171.8 (CO-Lys), 171.0 (CO-$Ala_{N\text{-}ter}$), 170.6 (CO-Met), 160.9 (CHO), 155.5 (OCONH), 77.3 (1C of ^{t}Bu), 52.5 (αC-Lys), 50.4 (αC-Met), 47.7 (αC-$Ala_{C\text{-}ter}$), 47.4 (αC-$Ala_{N\text{-}ter}$), 40.2 (εC-Lys), 32.2 (βC-Met), 31.5 (βC-Lys), 29.3 (γC-Met), 29.2 (δC-Lys), 28.3 (CH_3-^{t}Bu), 22.7 (γC-Lys), 18.1 (CH_3-$Ala_{C\text{-}ter}$), 17.2 (CH_3-$Ala_{N\text{-}ter}$), 14.6 (SCH_3); HRMS for $C_{23}H_{41}N_5O_8S$ $[M+Na]^+$ calcd.: 570.2568, found: 570.2571.

***N*-For-L-Met-L-Lys(Boc)-L-Ala-L-Phe-OMe (2.6b):** Following Method (b), tetrapeptide **2.6b** was obtained as a white solid (0.55 g, 82%), in one-pot without filtration of DCU, starting from **2.4i** (0.5 g, 1.05 mmol), *N*-hydroxysuccinimide (0.15 g, 1.26 mmol), DCC (0.28 g, 1.36 mmol) in THF (10 mL) and H-L-Phe-OMe·HCl (0.25 g, 1.15 mmol), $NaHCO_3$ (0.26 g, 3.15 mmol) in 1:1 THF/H_2O (30 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.29-8,24 (m, 2H, αN*H*-Met and αN*H*-Phe), 8.05-8.02 (m, 2H, C*H*O and αN*H*-Lys), 7.90 (d, J = 7.4 Hz, 1H, αN*H*-Ala), 7.30-7.19 (m, 5H, Ar-*H* Phe), 6.73 (t, J = 5.7 Hz, 1H, εN*H*-Lys), 4.49-4.44 (m, 1H, *αH*-Met), 4.32-4.27 (m, 1H, *αH*-Ala), 4.25-4.19 (m, 1H, *αH*-Lys), 3.57 (s, 3H, OCH_3), 3.05-2.94 (m, 2H, βCH_2-Phe), 2.92-2.82 (m, 2H, εCH_2-Lys), 2.43 (t, J = 7.8 Hz, 2H, γCH_2-Met), 2.02 (s, 3H, -SCH_3), 1.95-1.73 (m, 2H, βCH_2-Met), 1.65-1.44 (m, 2H, βCH_2-Lys), 1.40-1.24 (m, 11H, ^{t}Bu and δCH_2-Lys), 1.24-1.13 (m, 5H, CH_3-Ala and γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.2 (CO-Ala), 171.7 (CO-Phe), 171.0 (CO-Lys), 170.6 (CO-Met), 160.9 (CHO), 155.5 (OCONH), 136.9 (1C of Phe), 129.0 (Ar-C), 128.2 (Ar-C), 126.5 (Ar-C), 77.3 (1C of ^{t}Bu), 53.5 (αC-Phe), 52.4 (αC-Lys), 51.8 (OCH_3), 50.4 (αC-Met), 47.8 (αC-Ala), 40.2 (εC-Lys), 36.6 (βC-Phe), 32.2 (βC-Met), 31.5 (βC-Lys), 29.3 (γC-Met), 29.2 (δC-Lys), 28.2 (CH_3-^{t}Bu), 22.7 (γC-Lys), 18.2 (CH_3-Ala), 14.6 (SCH_3); HRMS for $C_{30}H_{47}N_5O_8S$ $[M+Na]^+$ calcd.: 660.3038, found: 660.3049.

***N*-For-L-Met-L-Lys(NH_2·TFA)-L-Ala-L-Phe-OMe (2.9):** Following a similar procedure as used for the synthesis of **2.8**, compound **2.6b** (248 mg, 0.388 mmol), thioanisole (0.046 mL, 0.388 mmol) and trifluoroacetic acid (1 mL) in CH_2Cl_2 (4 mL) afforded **2.9** as a colorless foam (253 mg, quantitative), which was used directly in the next step without any further purification.

***N*-For-L-Met-L-Lys(Boc)-L-Ala-L-Ala-L-Phe-OMe (2.7):** Following Method (b), pentapeptide **2.7** was obtained as a white solid (0.87 g, 81%), in two steps, starting from **2.6a** (0.83 g, 1.51 mmol), *N*-hydroxysuccinimide (0.21 g, 1.82 mmol), DCC (0.41 g, 1.97 mmol) in THF (20 mL) and H-L-Phe-OMe·HCl (0.36 g, 1.67 mmol), $NaHCO_3$ (0.38 g, 4.54 mmol) in 1:1 THF/H_2O (38 mL). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.28 (d, J = 8.3 Hz, 1H, αN*H*-Met), 8.23 (d, J = 7.6 Hz, 1H, αN*H*-Phe), 8.06 (d, J = 7.7 Hz, 1H, αN*H*-Lys), 8.02 (s, 1H, C*H*O), 7.96 (d, J = 7.2 Hz, 1H, αN*H*-Ala), 7.85 (d, J = 7.7 Hz, 1H, αN*H*-Ala), 7.30-7.19 (m, 5H, Ar-*H* Phe), 6.73 (t, J = 5.4 Hz, 1H, εN*H*-Lys), 4.49-4.40 (m, 2H, *αH*-Met and *αH*-Phe), 4.30-4.20 (unresolved m, 3H, 2 x *αH*-Ala and *αH*-Lys), 3.58 (s, 3H, OCH_3), 3.06-2.93 (m, 2H, βCH_2-Phe), 2.90-2.84 (m, 2H, εCH_2-Lys), 2.43 (t, J = 7.7 Hz, 2H, γCH_2-Met), 2.03 (s, 3H, -SCH_3), 1.95-1.73 (m, 2H, βCH_2-Met), 1.72-1.47 (m, 2H, βCH_2-Lys), 1.40-1.31 (m, 11H, ^{t}Bu and δCH_2-Lys), 1.24-1.12 (m, 8H, 2 x CH_3-Ala and γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.2 (CO-$Ala_{C\text{-}ter}$), 171.7 (CO-Phe), 171.5 (CO-$Ala_{N\text{-}ter}$), 171.2 (CO-Lys), 170.6 (CO-Met), 161.0 (CHO), 155.5 (OCONH), 137.0 (1C of Phe), 129.0 (Ar-C), 128.2 (Ar-C), 126.5 (Ar-C), 77.3 (1C of

^{t}Bu), 53.5 (αC-Phe), 52.5 (αC-Lys), 51.8 (OCH_3), 50.4 (αC-Met), 47.9 (αC-Ala), 47.9 (αC- $Ala_{C\text{-}ter}$), 47.7 (αC- $Ala_{N\text{-}ter}$), 40.2 (εC-Lys), 36.6 (βC-Phe), 32.2 (βC-Met), 31.5 (βC-Lys), 29.3 (γC-Met), 29.2 (δC-Lys), 28.3 (CH_3-^{t}Bu), 22.7 (γC-Lys), 18.2 (CH_3-Ala), 17.9 (CH_3-Ala), 14.6 (SCH_3); HRMS for $C_{33}H_{52}N_6O_9S$ $[M-H]^-$ calcd.: 707.3443, found: 707.3451.

***N*-For-L-Met-L-Lys(NH_2·TFA)-L-Ala-L-Ala-L-Phe-OMe (2.10):** Following a similar procedure as used for the synthesis of **2.8**, compound **2.7** (275 mg, 0.388 mmol), thioanisole (0.046 mL, 0.388 mmol) and trifluoroacetic acid (1 mL) in CH_2Cl_2 (4 mL) afforded **2.10** as a colorless foam (281 mg, quantitative), which was used directly in the next step without any further purification.

5'-*O*-(4-Monomethoxytrityl)-2'-deoxythymidine (2.14):[23] To a stirring solution of thymidine (2.50 g, 10.32 mmol) in dry DMF (26 mL) was added triethylamine (3.61 mL, 25.8 mmol), followed by DMAP (0.10 g, 0.825 mmol) and 4-monomethoxytrityl chloride (6.37 g, 20.64 mmol) and the resulting mixture was stirred at r.t. for 4 h. The reaction mixture was quenched with water (260 mL) and the aqueous layer was extracted with ethyl acetate (3 x 150 mL). The collected organic fractions were dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient Hexane/Ethyl acetate 2:1, v/v; 1:1, v/v; 1:2, v/v) to give **2.14** (5.14 g, 97%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.79 (br s, 1H, N*H*), 7.57 (d, *J* = 1.1 Hz, 1H, H-6), 7.42-7.39 (m, 4H, Ar-*H*), 7.33-7.22 (m, 8H, Ar-*H*), 6.86-6.83 (m, 2H, Ar-*H*), 6.42 (dd, *J* = 7.6, 6.1 Hz, 1H, H-1'), 4.59-4.57 (m, 1H, H-3'), 4.08-4.05 (m, 1H, H-4'), 3.79 (s, 3H, OC*H*$_3$), 3.49-3.35 (m, 2H, H-5' and H-5''), 2.44-2.29 (m, 2H, H-2' and H-2''), 1.47 (d, *J* = 1.1 Hz, 3H, C*H*$_3$); HRMS for $C_{30}H_{30}N_2O_6$ $[M+Na]^+$calcd.: 537.1996, found: 537.1998.

5'-*O*-(4-Monomethoxytrityl)-3'-*O*-(4-nitrophenyl carbonate)-2'-deoxythymidine (2.15): Compound **2.14** (1.00 g, 1.95 mmol) was coevaporated with dry pyridine (2 x 5 mL), redissolved in dry pyridine (5 ml) and to this solution was slowly added a suspension of 4-nitrophenyl chloroformate (0.432 g, 2.14 mmol) in a mixture of CH_2Cl_2/Pyridine (70:1, 5 mL) at 0 °C. The reaction mixture was then stirred overnight at r.t. After removal of all the volatiles under reduced pressure, the resulting residue was redissolved in CH_2Cl_2 (4 mL) and this solution was slowly added into 30 mL of vigorously stirring diethyl ether. The resulting precipitate was filtered off and the filtrate was poured into stirring hexane (100 mL) to give a white precipitate, which was filtered and dried to afford **2.15** (0.87 g, 66%) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.81 (br s, 1H, N*H*), 8.31-8.27 (m, 2H, Ar-*H*), 7.61 (s, 1H, H-6), 7.39- 7.26 (m, 14H, Ar-*H*), 6.87-6.84 (m, 2H, Ar-*H*), 6.39 (dd, *J* = 8.8, 5.6 Hz, 1H, H-1'), 5.51-5.45 (m, 1H, H-3'), 4.32 (br s, 1H, H-4'), 3.79 (s, 3H, OC*H*$_3$), 3.60-3.47 (m, 2H, H-5' and H-5''), 2.79-2.47 (m, 2H, H-2' and H-2''), 1.43 (s, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 163.6, 159.1, 155.2, 152.0, 150.5, 145.7, 143.7, 135.2, 134.6, 130.5, 128.5, 128.4, 128.3, 127.6, 125.5, 122.2, 121.8, 113.5 112.0, 84.4, 83.6, 80.5, 63.9, 55.4, 38.0, 11.8; HRMS for $C_{37}H_{33}N_3O_{10}$ $[M+Na]^+$calcd.: 702.2058, found: 702.2081.

5'-*O*-(4-Monomethoxytrityl)-*3'-O*-[*N*-For-Gly-L-Lys(ε-carbamate)-L-Phe-OMe]-2'-deoxythymidine (2.16): To a stirring solution of tripeptide **2.5d** (110 mg, 0.217 mmol) and Et_3N (0.105 mL, 0.868 mmol) in dry CH_2Cl_2 (5 mL) was added **2.15** (132 mg, 0.195 mmol) in dry CH_2Cl_2 at 0 °C and the stirring was continued at r.t. for 16 h. The reaction mixture was then diluted with CH_2Cl_2 and washed twice with 10% aq. $NaHCO_3$. The organic layer was dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/EtOH 99:1, v/v; 97:3, v/v; 94:6, v/v) to afford **2.16** as a white solid (160 mg, 79%). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.39 (br s,1H, N*H*), 8.40 (d, *J* = 7.4 Hz, 1H, N*H*-Phe), 8.22 (t, *J* = 5.4 Hz, 1H, N*H*-Gly), 8.06 (s, 1H, C*H*O), 7.99 (d, *J* = 8.2 Hz, 1H, N*H*-Lys),

7.53 (s, 1H, H-6), 7.53-7.19 (m, 18H, Ar-*H* and εN*H*-Lys), 6.23 (dd, *J* = 7.8, 6.5 Hz, 1H, H-1'), 5.24-5.23 (m, 1H, H-3'), 4.48-4.41 (m, 1H, αH-Phe), 4.31-4.25 (m, 1H, αH-Lys), 4.05 (br s, 1H, H-4'), 3.75 (s, 2H, CH_2-Gly), 3.56 (s, 3H, OCH_3), 3.39-3.20 (m, 2H, H-5' and H-5''), 2.99-2.94 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.47-2.25 (m, 1H, H-2'and H-2''), 1.77 (s, 3H, CH_3-Thy), 1.62-1.55 (m, 2H, βCH_2-Lys), 1.54-1.45 (m, 5H, δCH_2-Lys and CH_3-Thy), 1.37-1.24 (m, 2H, γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 171.7, 171.6, 168.1, 163.6, 161.4, 158.3, 155.2, 150.4, 144.1, 143.8, 137.1, 135.4, 134.6, 130.1, 129.0, 128.2, 127.9, 127.1, 126.5, 113.3, 109.9, 83.6, 83.2, 74.2, 63.9, 55.1, 63.6, 52.1, 51.7, 40.4, 36.8, 36.5, 31.9, 29.1, 22.4, 11.6; HRMS for $C_{50}H_{56}N_6O_{12}$ $[M+Na]^+$calcd.: 955.3848, found: 955.3813.

5'-Hydroxy-3'-*O*-[*N*-For-Gly-L-Lys(ε-carbamate)-L-Phe-OMe]-2'-deoxythymidine (2.17): A solution of **2.16** (147 mg, 0.157 mmol) in 80% AcOH (2 mL) was stirred at r.t. for 3.5 h. The volatiles were then removed under reduced pressure and coevaporated twice with toluene. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 100:0, v/v; 98:2, v/v; 96:5, v/v) to give **2.17** (85 mg, 82%) as a white solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.32 (br s,1H, N*H*), 8.38 (d, *J* = 7.3 Hz, 1H, N*H*-Phe), 8.19 (t, *J* = 5.1 Hz, 1H, N*H*-Gly), 8.06 (s, 1H, C*H*O), 7.96 (d, *J* = 7.9 Hz, 1H, N*H*-Lys), 7.73 (s, 1H, H-6), 7.31-7.19 (m, 5H, Ar-*H*), 6.17 (dd, *J* = 8.5, 5.9 Hz, 1H, H-1'), 5.19 (t, *J* = 4.7 Hz, 1H, εN*H*-Lys), 5.11-5.09 (m, 1H, H-3'), 4.48-4.40 (m, 1H, αH-Phe), 4.31-4.23 (m, 1H, αH-Lys), 3.93 (br s, 1H, H-4'), 3.75 (d, *J* = 5.6 Hz, 2H, αCH_2-Gly), 3.63-3.60 (m, 2H, H-5' and H-5''), 3.57 (s, 3H, OCH_3), 3.05-2.89 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.33-2.17 (m, 2H, H-2' and H-2''), 1.78 (s, 3H, CH_3-Thy), 1.59-1.47 (m, 2H, βCH_2-Lys), 1.46-1.34 (m, 2H, δCH_2-Lys), 1.28-1.15 (m, 2H, γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 171.8, 171.6, 168.1, 163.7, 161.4, 155.4, 150.5, 137.1, 135.8, 129.1, 128.3, 126.6, 109.7, 85.5, 83.7, 74.6, 61.5, 40.4, 36.8, 36.5, 31.9, 30.7, 29.1, 22.4, 12.3; HRMS for $C_{30}H_{40}N_6O_{11}$ $[M+H]^+$calcd.: 661.2827, found: 661.2828.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-Gly-L-Lys(ε-carbamate)-L-Phe-OMe]-2'-deoxythymidine (2.18): To a stirred suspension of **2.17** (210 mg, 0.317 mmol) in dry CH_2Cl_2 (5 mL) was added 0.45M tetrazole in ACN (3.53 mL, 1.59 mmol) followed by dibenzyldiisopropyl phosphoramidite (0.24 mL, 0.697 mmol) at 0 ºC and the reaction mixture was stirred at r.t. for 12 h. After cooling to -40 ºC, 35% H_2O_2 (0.14 mL, 1.59 mmol) was slowly added and the resulting mixture was stirred first at 0 ºC for 1 h and then at r.t. for 1 h. The mixture was cooled and quenched with 10% (w/v) aq. $NaHSO_3$ (6 mL). Athe aqueous layer was extracted with CH_2Cl_2 (3 x 30 mL), the combined organic fractions were dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 96:4, v/v) to give **2.18** (230 mg, 79%) as a colorless foam.

5'-*O*-(4-Monomethoxytrityl)-3'-*O*-benzoyl-2'-deoxythymidine (2.20): To a stirring solution of **2.14** (3.09 g, 6.01 mmol) in pyridine (25 mL), benzoyl chloride (0.80 mL, 6.91 mmol) was slowly added at 0 ºC and resulting mixture was stirred for 3 h at 0 ºC. The reaction mixture was then diluted with CH_2Cl_2 (200 mL) and washed with saturated aq. $NaHCO_3$ and brine. The organic layer was dried over Na_2SO_4, filtered, evaporated and coevaporated with toluene. The cude residue was purified by column chromatography on silica gel (gradient Hexane/Ethyl acetate 3:1, v/v; 2:1, v/v; 1:1, v/v) to give **2.20** (3.53 g, 95%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.56 (br s,1H, N*H*), 8.06-8.03 (m, 2H, Ar-*H*), 7.66 (s, 1H, H-6), 7.60-7.25 (m, 15H, Ar-*H*), 6.87-6.84 (m, 2H, Ar-*H*), 6.56 (dd, *J* = 8.7, 5.7 Hz, 1H, H-1'), 5.74-5.72 (m, 1H, H-3'), 4.30-4.29 (m, 1H, H-4'), 3.78 (s, 3H, OCH_3), 3.61-3.51 (m, 2H, H-5' and H-5''), 2.69-2.51 (m, 2H, H-2' and H-2''), 1.43 (s, 3H, CH_3); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 166.0, 164.0, 158.9, 150.8, 143.9, 143.8, 135.4, 134.8, 133.6, 130.5, 129.8,

129.4, 128.6, 128.5, 128.4, 128.1, 127.4, 113.4, 111.8, 87.5, 84.5, 84.2, 75.9, 63.9, 55.3, 38.2, 11.8; HRMS for $C_{37}H_{34}N_2O_7$ $[M+Na]^+$calcd.: 641.2258, found: 641.2257.

5'-Hydroxy-3'-*O*-benzoyl-2'-deoxythymidine (2.21):[52] A solution of **2.20** (3.35 g, 5.41 mmol) in 80% AcOH (40 mL) was stirred at r.t. for 3.5 h. The volatiles were then removed under reduced pressure and twice coevaporated with toluene. The residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 100:0, v/v; 98:2, v/v; 96:4, v/v) to give **2.21** (1.7 g, 91%) as a white solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.35 (br s,1H, N*H*), 8.03-8.00 (m, 2H, *o*-H of Ph), 7.79 (d, *J* = 1.1 Hz, 1H, H-6), 7.71-7.66 (m, 1H, *p*-H of Ph), 7.58-7.53 (m, 2H, *m*-H of Ph), 6.30 (app t, *J* = 7.1 Hz, 1H, H-1'), 5.49-5.48 (m, 1H, H-3'), 5.26 (br s, 1H, O*H*), 4.18-4.15 (m, 1H, H-4'), 3.71 (br s, 2H, H-5' and H-5''), 2.43-2.39 (m, 2H, H-2' and H-2''), 1.80 (d, *J* = 1.0 Hz, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 165.2 (COPh), 163.7 (C-4), 150.5 (C-2), 135.9 (C-6), 133.6 (*p*-C of Ph), 129.4 (1C of Ph), 129.3 (*o*-C of Ph), 128.8 (*m*-C of Ph), 109.8 (C-5), 84.6 (C-1'), 83.8 (C-4'), 75.7 (C-3'), 61.4 (C-5'), 36.7 (C-2'), 12.3 (CH_3); HRMS for $C_{37}H_{34}N_2O_7$ $[M+H]^+$calcd.: 347.1237, found: 347.1238.

5'-*O*-(Dibenzylphosphate)-3'-*O*-benzoyl-2'-deoxythymidine (2.22): To a stirring suspension of **2.21** (1.6 g, 4.62 mmol) in dry CH_2Cl_2 (20 mL), 0.45M tetrazole in ACN (51.33 mL, 23.10 mmol) was added followed by dibenzyldiisopropyl phosphoramidite (3.47 mL, 10.16 mmol) at 0 ºC and the reaction mixture was stirred at r.t. for 12 h. After cooling to -40 ºC, 35% H_2O_2 (1.97 mL, 23.10 mmol) was slowly added and the resulting mixture was stirred at 0 ºC for 1 h and an additional 1 h at r.t. Then the reaction mixture was again cooled and finally quenched with 10% (w/v) aq. $NaHSO_3$ (30 mL). The aqueous layer was extracted with CH_2Cl_2 (3 x 100 mL), collected organic fractions were dried over Na_2SO_4, concentrated under reduced pressure and crude residue was purified by column chromatography on silica gel (gradient Hexane/Ethyl acetate 3:1, v/v; 2:1, v/v; 1:1, v/v) to give **2.22** (2.59 g, 92%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.38 (br s,1H, N*H*), 8.05-8.02 (m, 2H, Ar-*H*), 7.63-7.58 (m, 1H, Ar-*H*), 7.49-7.44 (m, 3H, H-6 and Ar-H), 7.35 (br s, 10H, Ar-*H*), 6.46 (dd, *J* = 9.0, 5.4 Hz, 1H, H-1'), 5.37-5.35 (m, 1H, H-3'), 5.15-5.03 (m, 4H, OC*H*$_2$Ph), 4.38-4.34 (m, 1H, H-4'), 4.25 (br s, 2H, H-5' and H-5''), 2.48-2.42 (m, 2H, H-2' and H-2''), 1.86 (s, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 166.0 (COPh), 163.8 (C-4), 150.7 (C-2), 135.5 (d, $^3J_{C,P}$ = 6.0 Hz, 1C of OCH_2Ph),135.0 (C-6), 133.7 (Ar-C), 129.8 (Ar-C), 129.1 (Ar-C), 129.0 (Ar-C), 128.8 (Ar-C), 128.6 (Ar-C), 128.2 (Ar-C), 111.9 (C-5), 84.6 (C-1'), 83.0 (d, $^3J_{C,P}$ = 8.3 Hz, C-4'), 75.3 (C-3'), 69.9 (app t, $^2J_{C,P}$ = 5.3 Hz, OCH_2), 67.3 (d, $^2J_{C,P}$ =5.7 Hz, C-5'), 37.3 (C-2'), 12.4 (CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.6; HRMS for $C_{37}H_{34}N_2O_7$ $[M+H]^+$calcd.: 607.1840, found: 607.1840.

5'-*O*-(Dibenzylphosphate)-2'-deoxythymidine (2.23): A solution of **2.22** (2.52 g, 4.15 mmol) in 7N NH_3 in MeOH (30 mL) was stirred at 0 ºC for 3 h and then at r.t. for 24 h. The volatiles were removed under reduced pressure and the crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 97:3, v/v) to give **2.23** (1.98 g, 95%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.41 (br s,1H, N*H*), 7.34- 7.33 (m, 11H, H-6 and Ar-*H*), 6.29 (app t, *J* = 6.7 Hz, 1H, H-1'), 5.11-4.99 (m, 4H, OC*H*$_2$Ph), 4.37-4.33 (m, 1H, H-3'), 4.22-4.14 (m, 2H, H-5' and H-5''), 4.04-4.00 (m, 2H, H-4'), 2.35-2.27 (m, 1H, H-2'), 2.03-1.96 (m, 1H, H-2''), 1.81 (d, *J* = 1.0 Hz, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 164.0 (C-4), 150.7 (C-2), 149.8 (C1 of $PhNO_2$), 145.36 (p-C of $PhNO_2$), 135.7 (C-6), 135.4 (d, $^3J_{C,P}$ = 7.3 Hz, 1C of OCH_2Ph), 129.1 (Ar-C), 128.9 (Ar-C), 128.2 (Ar-C), 111.5 (C-5), 85.0 (C-1'), 84.8 (d, $^3J_{C,P}$ = 7.5 Hz, C-4'), 71.1 (C-3'), 70.1-70.0 (2d, $^2J_{C,P}$ = 5.6 Hz, 2 x OCH_2Ph), 66.9 (d, $^2J_{C,P}$ = 5.7 Hz, C-5'), 40.2 (C-2'), 12.5 (CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.5; HRMS for $C_{24}H_{27}N_2O_8P$ $[M+H]^+$calcd.: 503.1578, found: 503.1580.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(4-nitrophenyl carbonate)-2'-deoxythymidine (2.24): Compound **2.23** (1.12 g, 2.23 mmol) was coevaporated with dry pyridine (2 x 5 mL), redissolved in dry pyridine (7 ml) and then a suspension of 4-nitrophenyl chloroformate (0.539 g, 2.67 mmol) in a mixture of CH_2Cl_2/Pyridine (70:1, 7.1 mL) was slowly added at 0 ºC . The reaction mixture was then stirred at 0 ºC for 2 h and at r.t. for 24 h. The volatiles were removed under reduced pressure and the residue was redissolved in CH_2Cl_2 (150 mL) and washed with saturated aq. $NaHCO_3$ (2 x 70 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient Hexane/Ethyl acetate 2:1, v/v; 1:1, v/v; 1:3, v/v) to afford **2.24** (0.99 g, 67%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.43 (br s,1H, N*H*), 8.32-8.29 (m, 2H, Ar-*H*), 7.41- 7.26 (m, 13H, H-6 and Ar-*H*), 6.39 (dd, *J* = 9.3, 5.3 Hz, 1H, H-1'), 5.17-5.15 (m, 1H, H-3'), 5.12-5.00 (m, 4H, OC*H*$_2$Ph), 4.28-4.27 (m, 1H, H-4'), 4.26-4.20 (m, 2H, H-5' and H-5''), 2.48-2.41 (m, 1H, H-2'), 2.04-1.93 (m, 1H, H-2''), 1.86 (d, *J* = 1.0 Hz, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 162.9 (C-4), 154.7 ($COPhNO_2$), 151.5 (C-2), 149.8 (C1 of $PhNO_2$), 145.36 (p-C of $PhNO_2$), 135.0 (d, $^3J_{C,P}$ = 5.8 Hz, 1C of OCH_2Ph), 134.4 (C-6), 128.7 (Ar-C), 128.6 (Ar-C), 128.5 (Ar-C), 128.4 (Ar-C), 127.8 (Ar-C), 125.1 (Ar-C), 121.3 (Ar-C), 111.6 (C-5), 84.1 (C-1'), 81.9 (d, $^3J_{C,P}$ = 8.2 Hz, C-4'), 79.1 (C-3'), 69.6 (app t, $^2J_{C,P}$ = 4.9 Hz, OCH_2Ph), 66.6 (d, $^2J_{C,P}$ =5.5 Hz, C-5'), 36.7 (C-2'), 12.0 (CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.5; HRMS for $C_{31}H_{30}N_3O_{12}P$ $[M+H]^+$calcd.: 668.1640, found: 668.1638.

General procedure for the synthesis of 3'-carbamoyl peptide conjugates of dTMP (2.18 and 2.25-2.32): To a stirring solution of peptide-NH_2·TFA salt (1 eq.) and Et_3N (4 eq.) in dry CH_2Cl_2 was added a solution of compound **2.24** in dry CH_2Cl_2 at 0 ºC and the stirring was continued at r.t. for 24 h. The reaction mixture was then diluted with CH_2Cl_2 and washed twice with 20% aq. $NaHCO_3$. The organic layer was dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 96:4, v/v) to afford the desired 3'-carbamoyl PNCs.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-Gly-L-β-amino-Ala-(β-carbamate)-L-Phe-OMe]-2'-deoxythymidine (2.25): Compound **2.25** was prepared according to the general procedure starting from **2.24** (240 mg , 0.365 mmol), tripeptide **2.5c** (179 mg, 0.510 mmol) and Et_3N (0.05 mL, 0.539 mmol) in dry CH_2Cl_2 (15 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 92:5, v/v) to give conjugate **2.25** (259 mg, 82%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.93 (br s, 1H, N*H*-Thy), 8.18 (s, 1H, C*H*O), 7.65 (d, *J* = 7.2 Hz, 1H, N*H*-Phe), 7.65 (d, *J* = 7.2 Hz, 1H, N*H*-β-amino-Ala), 7.39 (s, 1H, H-6), 7.32 (br s, 10H, Ar-*H* of OBn), 7.26-7.09 (m, 5H, Ar-*H* of Phe), 6.39 (t, *J* = 6.3 Hz, 1H, N*H*-Gly), 6.25 (dd, *J* = 8.3, 5.2 Hz, 1H, H-1'), 5.07-4.98 (m, 5H, H-3' and 2 x OC*H*$_2$Ph), 4.83-4.76 (m, 1H, *αH*-Phe), 4.75-4.68 (m, 1H, *αH*-β-amino-Ala), 4.28-4.17 (m, 2H, H-5' and H-5''), 4.14 (br s, 1H, H-4'), 4.04-3.88 (m, 2H, C*H*$_2$-Gly), 3.67 (s, 3H, OC*H*$_3$), 3.58-3.33 (m, 2H, βC*H*$_2$-β-amino-Ala), 3.16-2.99 (m, 2H, βC*H*$_2$-Phe), 2.38-2.14 (m, 2H, H-2' and H-2''), 1.80 (s, 3H, C*H*$_3$-Thy); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.0 (CO-Phe), 169.7 (CO-β-amino-Ala), 169.4 (CO-Gly), 164.0 (C-4), 162.4 (CHO), 156.5 (OCONH), 151.1 (C-2), 136.1 (C-6), 135.5 (d, $^3J_{C,P}$ = 5.5 Hz, 1C of OCH_2Ph), 135.1 (1C Ph-Phe), 129.3 (Ar-C), 129.0 (Ar-C), 128.9 (Ar-C), 128.6 (Ar-C), 128.2 (Ar-C), 127.2 (Ar-C), 112.0 (C-5), 84.6 (C-1'), 82.8 (d, $^3J_{C,P}$ = 7.7 Hz, C-4'), 75.5 (C-3'), 69.9 (app t, $^2J_{C,P}$ = 5.0 Hz, 2 x OCH_2Ph), 67.4 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 53.9 (αC-Phe), 53.6 (αC-β-amino-Ala), 52.6 (OCH_3), 43.3 (βC-β-amino-Ala), 41.7 (αC-Gly), 37.8 (C-2'), 37.2 (βC-Phe), 12.5 (CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.7; HRMS for $C_{41}H_{47}N_6O_{14}P$ $[M+H]^+$calcd.: 879.2960, found: 879.2966.

5'-*O*-(Dibenzylphosphate)-*3'-O*-[*N*-For-Gly-L-Lys(ε-carbamate)-L-Phe-OMe]-2'-deoxythymidine (2.18): Compound **2.18** was prepared according to the general procedure starting from **2.24** (300 mg , 0.449 mmol), tripeptide **2.5d** (273 mg, 0.539 mmol) and Et_3N (0.25 mL, 1.80 mmol) in dry CH_2Cl_2 (15 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:2, v/v; 92:4, v/v) to give conjugate **2.18** (322 mg, 78%) as a colorless foam. ^{1}H NMR (300 MHz, CD_3OD) δ = 8.15 (s, 1H, C*H*O), 7.46 (s, 1H, H-6), 7.36 (br s, 10H, Ar-*H* of OBn), 7.29-7.17 (m, 5H, Ar-*H* of Phe), 6.23 (dd, *J* = 8.6, 5.8 Hz, 1H, H-1'), 5.12-5.08 (m, 5H, H-3' and 2 x OCH_2Ph), 4.67-4.63 (m, 1H, *αH*-Phe), 4.38-4.33 (m, 1H, *αH*-Lys), 4.30-4.26 (m, 2H, H-5' and H-5''), 4.17-4.16 (m, 1H, H-4'), 3.90 (s, 2H, CH_2-Gly), 3.67 (s, 3H, OCH_3), 3.16-2.96 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.33-2.27 (m, 1H, H-2'), 2.12-2.02 (m, 1H, H-2''), 1.77 (s, 3H, CH_3-Thy), 1.71-1.58 (m, 2H, βCH_2-Lys), 1.51-1.47 (m, 2H, δCH_2-Lys), 1.33-1.29 (m, 2H, γCH_2-Lys); ^{13}C NMR (75 MHz, CD_3OD) δ = 173.9 (CO-Phe), 173.3 (CO-Lys), 170.7 (CO-Gly), 166.1 (C-4), 164.2 (CHO), 157.8 (OCONH), 152.2 (C-2), 138.1 (1C Ph-Phe), 137.1 (C-6), 137.0 (d, $^3J_{C,P}$ = 6.1 Hz, 1C of OCH_2Ph), 130.3 (Ar-C), 129.9 (Ar-C), 129.8 (Ar-C), 129.5 (Ar-C), 129.2 (Ar-C), 127.9 (Ar-C), 112.1 (C-5), 86.2 (C-1'), 84.4 (d, $^3J_{C,P}$ = 7.7 Hz, C-4'), 75.7 (C-3'), 71.2-71.1 (2d, $^2J_{C,P}$ = 5.7 Hz, 2 x OCH_2Ph), 68.7 (d, $^2J_{C,P}$ = 5.8 Hz, C-5'), 55.2 (αC-Phe), 54.4 (αC-Lys), 52.7 (OCH_3), 42.0 (αC-Gly), 41.5 (εC-Lys), 38.3 (C-2'), 38.1 (βC-Phe), 32.8 (βC-Lys), 30.3 (δC-Lys), 23.8 (γC-Lys), 12.5 (CH_3); ^{31}P NMR (121 MHz, CD_3OD) δ = -1.2; HRMS for $C_{44}H_{53}N_6O_{14}P$ $[M-H]^-$calcd.: 919.3284, found: 919.3287.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Phe-OMe]-2'-deoxythymidine (2.26): Compound **2.26** was prepared according to the general procedure starting from **2.24** (246.2 mg, 0.367 mmol), tripeptide **2.5g** (225.4 mg, 0.3882 mmol) and Et_3N (0.22 mL, 1.55 mmol) in dry CH_2Cl_2 (20 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 92:5, v/v) to give conjugate **2.26** (330 mg, 85%) as a colorless foam. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.36 (br s, 1H, N*H*-Thy), 8.33 (d, *J* = 7.5 Hz, 1H, N*H*-Phe), 8.27 (d, *J* = 8.4 Hz, 1H, N*H*-Met), 8.01-8.00 (m, 2H, C*H*O and αN*H*-Lys), 7.50 (s, 1H, H-6), 7.36-7.34 (m, 11H, Ar-*H* of OBn and εN*H*-Lys), 7.26-7.19 (m, 5H, Ar-*H* of Phe), 6.19 (app t, *J* = 7.3 Hz, 1H, H-1'), 5.07-5.04 (m, 5H, H-3' and 2 x OCH_2Ph), 4.50-4.39 (m, 2H, *αH*-Met and *αH*-Phe), 4.26-4.20 (unresolved m, 3H, *αH*-Lys, H-5' and H-5''), 4.10 (br s, 1H, H-4') 3.56 (s, 3H, OCH_3), 3.05-2.89 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.39 (t, *J* = 7.8 Hz, 2H, γCH_2-Met), 2.23-2.18 (m, 2H, H-2' and H-2''), 2.00 (s, 3H, SCH_3), 1.91-1.72 (m, 2H, βCH_2-Met), 1.69 (s, 3H, CH_3-Thy), 1.63-1.45 (m, 2H, βCH_2-Lys), 1.40-1.36 (m, 2H, δCH_2-Lys), 1.28-1.19 (m, 2H, γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 171.7 (CO-Phe), 171.5 (CO-Lys), 170.4 (CO-Met), 163.5 (C-4), 160.9 (CHO), 155.2 (OCONH), 150.4 (C-2), 137.0 (1C of Phe), 135.9 (d, $^3J_{C,P}$ = 6.8 Hz, 1C of OCH_2Ph), 135.4 (C-6), 129.0 (Ar-C), 128.5 (Ar-C), 128.4 (Ar-C), 128.2 (Ar-C), 127.8 (Ar-C), 126.5 (Ar-C), 110.0 (C-5), 83.9 (C-1'), 82.2 (d, $^3J_{C,P}$ = 7.4 Hz, C-4'), 73.6 (C-3'), 68.7 (2d, $^2J_{C,P}$ = 5.5 Hz, 2 x OCH_2Ph), 67.0 (d, $^2J_{C,P}$ = 5.7 Hz, C-5'), 53.4 (αC-Phe), 52.2 (αC-Lys), 51.8 (OCH_3), 50.3 (αC-Met), 40.2 (εC-Lys), 36.5 (βC-Phe), 36.0 (C-2'), 32.2 (βC-Met), 31.7 (βC-Lys), 29.3 (γC-Met), 29.0 (δC-Lys), 22.5 (γC-Lys), 14.6 (-SCH_3), 12.0 (CH_3-Thy); ^{31}P NMR (121 MHz, DMSO-d_6) δ = -0.9; HRMS for $C_{47}H_{59}N_6O_{14}PS$ $[M-H]^-$calcd.: 993.3474, found: 993.3478.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Ala-OMe]-2'-deoxythymidine (2.27): Compound **2.27** was prepared according to the general procedure starting from **2.24** (226.6 mg, 0.339 mmol), tripeptide **2.5e** (205.5 mg, 0.407 mmol) and Et_3N (0.19 mL, 1.36 mmol) in dry CH_2Cl_2 (15 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 92:5, v/v) to give conjugate **2.27** (224.5 mg, 72%) as a colorless foam. ^{1}H NMR (500 MHz, DMSO-d_6) δ = 11.3 (br s, 1H, N*H*-Thy), 8.31 (d, *J* = 6.8 Hz, 1H,

N*H*-Ala), 8.27 (d, *J* = 8.1 Hz, 1H, N*H*-Met), 8.03-8.01 (m, 2H, C*H*O and αN*H*-Lys), 7.50 (s, 1H, H-6), 7.8a-7.35 (m, 11H, Ar-*H* of OBn and εN*H*-Lys), 6.19 (app t, *J* = 7.3 Hz, 1H, H-1'), 5.08-5.05 (m, 5H, H-3' and 2 x OCH_2Ph), 4.46-4.41 (m, 1H, *αH*-Met), 4.27-4.22 (unresolved m, 4H, *αH*-Lys, *αH*-Ala, H-5' and H-5''), 4.11 (br s, 1H, H-4'), 3.61 (s, 3H, OCH_3), 2.99-2.95 (m, 2H, εCH_2-Lys), 2.42 (t, *J* = 7.9 Hz, 2H, γCH_2-Met), 2.22-2.19 (m, 2H, H-2' and H-2''), 2.02 (s, 3H, SCH_3), 1.91-1.72 (m, 2H, βCH_2-Met), 1.69 (s, 3H, CH_3-Thy), 1.67-1.48 (m, 2H, βCH_2-Lys), 1.43-1.37 (m, 2H, δCH_2-Lys), 1.33-1.22 (m, 5H, γCH_2-Lys and CH_3-Ala); ^{13}C NMR (125 MHz, DMSO-d_6) δ = 172.9 (CO-Ala), 171.4 (CO-Lys), 170.5 (CO-Met), 163.6 (C-4), 161.0 (CHO), 155.2 (OCONH), 150.4 (C-2), 135.9 (d, $^3J_{C,P}$ = 6.9 Hz, 1C of OCH$_2$Ph), 135.4 (C-6), 128.5 (Ar-C), 128.4 (Ar-C), 127.8 (Ar-C), 110.0 (C-5), 83.9 (C-1'), 82.2 (d, $^3J_{C,P}$ = 7.2 Hz, C-4'), 73.6 (C-3'), 68.7 (d, $^2J_{C,P}$ = 5.4 Hz, 2 x OCH$_2$Ph), 67.1 (d, $^2J_{C,P}$ = 5.9 Hz, C-5'), 52.1 (αC-Lys), 51.8 (OCH_3), 50.4 (αC-Met), 47.5 (αC-Ala), 40.2 (εC-Lys), 36.0 (C-2'), 32.2 (βC-Met), 31.6 (βC-Lys), 29.3 (γC-Met), 29.1 (δC-Lys), 22.5 (γC-Lys), 16.8 (CH$_3$-Ala), 14.6 (-SCH$_3$), 12.0 (CH$_3$-Thy); ^{31}P NMR (202 MHz, DMSO-d_6) δ = -0.9; HRMS for $C_{41}H_{55}N_6O_{14}PS$ [M-H]$^-$calcd.: 917.3161, found: 917.3145.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Lys(Cbz)-OMe]-2'-deoxythymidine (2.28): Compound **2.28** was prepared according to the general procedure starting from **2.24** (326.3 mg, 0.489 mmol), tripeptide **2.5f** (408.1 mg, 0.587 mmol) and Et$_3$N (0.33 mL, 2.35 mmol) in dry CH$_2$Cl$_2$ (15 ml). The crude residue was purified by column chromatography on silica gel (gradient CH$_2$Cl$_2$/MeOH 99:1, v/v; 97:2, v/v; 92:4, v/v) to give conjugate **2.28** (347 mg, 64%) as a colorless foam. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.53 (br s, 1H, N*H*-Thy), 8.31-8.25 (m, 2H, 2 x N*H*), 8.03-8.01 (m, 2H, C*H*O and N*H*), 7.50 (s, 1H, H-6), 7.36-7.19 (m, 15H, Ar-*H* of OBn), 7.23-7.19 (m, 1H, N*H*), 6.18 (app t, *J* = 7.2 Hz, 1H, H-1'), 5.07-5.04 (m, 5H, H-3' and 2 x POCH_2Ph-OBn), 4.99 (s, 2H, OCH_2-Cbz), 4.47-4.39 (m, 1H, *αH*-Met), 4.30-4.15 (m, 4H, 2 x *αH*-Lys, H-5' and H-5''), 4.09 (br s, 1H, H-4'), 3.61 (s, 3H, OCH_3), 2.99-2.92 (m, 4H, 2 x εCH_2-Lys), 2.41 (t, *J* = 7.8 Hz, 2H, γCH_2-Met), 2.22-2.16 (m, 2H, H-2' and H-2''), 2.01 (s, 3H, SCH_3), 1.90-1.76 (m, 2H, βCH_2-Met), 1.69 (s, 3H, CH_3-Thy), 1.67-1.49 (m, 4H, 2 x βCH_2-Lys), 1.42-1.35 (m, 4H, 2 x δCH_2-Lys), 1.33-1.17 (m, 4H, 2 x γCH_2-Lys); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.4 (CO-Lys$_{C\text{-}Ter}$), 171.6 (CO-Met), 170.4 (CO-Lys$_{N\text{-}Ter}$), 163.5 (C-4), 161.0 (CHO), 155.2 (2 x OCONH), 150.4 (C-2), 137.2 (C-6), 135.9 (d, $^3J_{C,P}$ = 6.7 Hz, 1C of OCH$_2$Ph), 135.4 (1C of Ph-Cbz), 128.5 (Ar-C), 128.4 (Ar-C), 128.3 (Ar-C), 127.8 (Ar-C), 127.7 (Ar-C), 127.6 (Ar-C), 110.0 (C-5), 83.9 (C-1'), 82.2 (d, $^3J_{C,P}$ = 8.2 Hz, C-4'), 73.6 (C-3'), 68.8-68.7 (3 x OCH$_2$Ph), 65.1 (d, $^2J_{C,P}$ = 3.7 Hz, C-5'), 52.3 (αC-Lys), 52.6 (αC-Lys and OCH$_3$), 50.4 (αC-Met), 39.5 (2 x εC-Lys merged with DMSO-d_6), 36.0 (C-2'), 32.1 (βC-Met), 31.8 (2 x βC-Lys), 30.5 (γC-Met), 29.3 (δC-Lys), 28.9 (δC-Lys), 22.5 (2 x γC-Lys), 14.5 (SCH$_3$), 12.0 (CH$_3$-Thy); ^{31}P NMR (121 MHz, DMSO-d_6) δ = -0.9; HRMS for $C_{52}H_{68}N_7O_{16}PS$ [M-H]$^-$calcd.: 1108.4108, found: 1108.4100.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-Boc-L-Ala-L-Lys(ε-carbamate)-L-Ala-OMe]-2'-deoxythymidine (2.29): Compound **2.29** was prepared according to the general procedure starting from **2.24** (505 mg, 0.756 mmol), tripeptide **2.13** (350 mg, 0.89 mmol) and Et$_3$N (0.16 mL, 1.13 mmol) in dry CH$_2$Cl$_2$ (20 ml). The crude residue was purified by column chromatography on silica gel (gradient CH$_2$Cl$_2$/MeOH 99:1, v/v; 97:2, v/v; 92:4, v/v) to give conjugate **2.29** (619 mg, 88%) as a colorless foam. ^{1}H NMR (500 MHz, CDCl$_3$) δ = 9.41 (br s, 1H, N*H*-Thy), 7.44 (s, 1H, H-6), 7.33 (br s, 10H, Ar-*H* of OBn), 7.06 (br s, 2H, N*H*-Ala$_{C\text{-}Ter}$ and N*H*-Lys), 6.29 (dd, *J* = 9.1, 5.3 Hz, 1H, H-1'), 5.62 (br s, 1H, ε-N*H*-Lys), 5.39 (br s, 1H, N*H*-Ala$_{N\text{-}Ter}$), 5.10-5.01 (m, 5H, H-3' and 2 x OCH_2Ph), 4.55-4.48 (m, 2H, *αH*-Ala$_{C\text{-}Ter}$ and *αH*-Lys), 4.29-4.26 (m, 2H, H-5'), 4.21-4.19 (m, 2H, *αH*-Ala$_{N\text{-}Ter}$ and H-5''), 4.16 (br s, 1H, H-4'), 3.74 (s, 3H, OCH_3), 3.58-3.33 (m, 2H, εCH_2-Lys), 2.31-2.25 (m, 2H, H-2'), 1.93-1.86 (m, 2H, H-2'' and βCH_2-Lys), 1.84 (s, 3H, CH_3-Thy), 1.71-1.63 (m, 1H, βCH_2-

Lys), 1.57-1.48 (m, 1H, δCH_2-Lys), 1.42 (s, 9H, ^{t}Bu), 1.40-1.34 (m, 8H, γCH_2-Lys, 2 x CH_3-Ala); ^{13}C NMR (125 MHz, $CDCl_3$) δ = 173.4 (CO-$Ala_{C\text{-}Ter}$), 173.2 (CO-$Ala_{N\text{-}Ter}$), 171.3 (CO-Lys), 163.9 (C-4), 155.7 (2 x OCONH), 150.8 (C-2), 135.5 (d, $^3J_{C,P}$ = 5.6 Hz, 1C of OCH_2Ph), 135.2 (C-6), 129.0 (Ar-C), 128.8 (Ar-C), 128.2 (Ar-C), 111.9 (C-5), 84.6 (C-1'), 83.1 (d, $^3J_{C,P}$ = 8.0 Hz, C-4'), 80.2 (1C ^{t}Bu), 75.0 (C-3'), 69.9-69.8 (2d, $^2J_{C,P}$ = 5.4 Hz, 2 x OCH_2Ph), 67.4 (d, $^2J_{C,P}$ = 5.8 Hz, C-5'), 52.8 (αC- $Ala_{C\text{-}Ter}$), 52.6 (OCH_3), 48.2 (αC-Lys and αC-$Ala_{N\text{-}Ter}$), 40.6 (εC-Lys), 37.4 (C-2'), 29.1 (βC-Lys), 28.4 (δC-Lys and CH_3-^{t}Bu), 22.4 (γC-Lys), 18.0 (2 x CH_3-Ala), 12.4 (CH_3-Thy); ^{31}P NMR (202 MHz, $CDCl_3$) δ = -0.7; HRMS for $C_{43}H_{59}N_6O_{15}P$ $[M\text{-}H]^-$calcd.: 929.3703, found: 929.3715.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)OMe]-2'-deoxythymidine (2.30): Compound **2.30** was prepared according to the general procedure starting from **2.24** (246.2 mg, 0.367 mmol), dipeptide **2.8** (168.3 mg, 0.3882 mmol) and Et_3N (0.22 mL, 1.55 mmol) in dry CH_2Cl_2 (15 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 92:4, v/v) to give conjugate **2.30** (276.5 mg, 84%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 10.30 (br s, 1H, N*H*-Thy), 8.20 (s, 1H, C*H*O), 7.56 (d, *J* = 7.9 Hz, 1H, N*H*-Met), 7.46 (s, 1H, H-6), 7.37-7.34 (m, 11H, Ar-*H* of OBn and αN*H*-Lys), 6.30 (dd, *J* = 9.0, 5.3 Hz, 1H, H-1'), 5.95 (t, *J* = 5.5 Hz, εN*H*-Lys), 5.12-5.04 (m, 5H, H-3' and 2 x OCH_2Ph), 4.76-4.69 (m, 1H, *αH*-Met), 4.61-4.54 (m, 1H, *αH*-Lys), 4.27-4.22 (unresolved m, 3H, H-4', H-5' and H-5''), 3.73 (s, 3H, OCH_3), 3.18-3.11 (m, 2H, εCH_2-Lys), 2.57 (t, *J* = 7.2 Hz, 2H, γCH_2-Met), 2.30-2.24 (m, 1H, H-2'), 2.17-1.98 (m, 5H, SCH_3 and βCH_2-Met), 1.92-1.65 (m, 6H, CH_3-Thy, H-2'' and βCH_2-Lys), 1.58-1.48 (m, 2H, δCH_2-Lys), 1.42-1.32 (m, 2H, γCH_2-Lys); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.5 (CO-Lys), 171.2 (CO-Met), 164.1 (C-4), 161.7 (CHO), 155.6 (OCONH), 151.1 (C-2), 135.4 (d, $^3J_{C,P}$ = 6.2 Hz, 1C of OCH_2Ph), 135.1 (C-6), 128.8 (Ar-C), 128.7 (Ar-C), 128.0 (Ar-C), 111.9 (C-5), 84.4 (C-1'), 83.0 (d, $^3J_{C,P}$ = 8.0 Hz, C-4'), 74.7 (C-3'), 69.7 (2d, $^2J_{C,P}$ = 5.5 Hz, 2 x OCH_2Ph), 67.3 (d, $^2J_{C,P}$ = 5.4 Hz, C-5'), 52.4 (OCH_3), 52.0 (αC-Lys), 51.0 (αC-Met), 40.4 (εC-Lys), 37.3 (C-2'), 31.5 (βC-Met), 29.9 (βC-Lys and γC-Met), 28.8 (δC-Lys), 22.4 (γC-Lys), 15.2 (-SCH_3), 12.3 (CH_3-Thy); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.8; HRMS for $C_{31}H_{44}N_5O_{13}PS$ $[M\text{-}H]^-$calcd.: 756.2320, found: 756.2347.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Ala-L-Phe-OMe]-2'-deoxythymidine (2.31): Compound **2.31** was prepared according to the general procedure starting from **2.24** (246.2 mg, 0.367 mmol), tetrapeptide **2.9** (253 mg, 0.3882 mmol) and Et_3N (0.22 mL, 1.55 mmol) in dry CH_2Cl_2 (20 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 92:5, v/v) to give conjugate **2.31** (339 mg, 82%) as a colorless foam. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.36 (br s, 1H, N*H*-Thy), 8.30-8.25 (m, 2H, N*H*-Phe and N*H*-Met), 8.05 (d, *J* = 8.0 Hz, 1H, αN*H*-Lys), 8.02 (s, 1H, C*H*O), 7.91 (d, *J* = 7.5 Hz, 1H, αN*H*-Ala), 7.50 (s, 1H, H-6), 7.36-7.32 (m, 11H, Ar-*H* of OBn and εN*H*-Lys), 7.29-7.19 (m, 5H, Ar-*H* of Phe), 6.19 (app t, *J* = 7.3 Hz, 1H, H-1'), 5.07-5.04 (m, 5H, H-3' and 2 x OCH_2Ph), 4.49-4.40 (m, 2H, *αH*-Met and *αH*-Phe), 4.33-4.18 (unresolved m, 4H, *αH*-Ala, *αH*-Lys, H-5' and H-5''), 4.12-4.08 (m, 1H, H-4') 3.57 (s, 3H, OCH_3), 3.05-2.90 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.43 (t, *J* = 7.8 Hz, 2H, γCH_2-Met), 2.23-2.17 (m, 2H, H-2' and H-2''), 2.02 (s, 3H, SCH_3), 1.95-1.75 (m, 2H, βCH_2-Met), 1.70 (s, 3H, CH_3-Thy), 1.64-1.45 (m, 2H, βCH_2-Lys), 1.41-1.34 (m, 2H, δCH_2-Lys), 1.29-1.22 (m, 2H, γCH_2-Lys), 1.17 (d, *J* = 6.9 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.2 (CO-Ala), 171.7 (CO-Phe), 171.0 (CO-Lys), 170.6 (CO-Met), 163.6 (C-4), 161.0 (CHO), 155.2 (OCONH), 150.4 (C-2), 136.9 (1C of Phe), 135.9 (d, $^3J_{C,P}$ = 6.9 Hz, 1C of OCH_2Ph), 135.4 (C-6), 129.0 (Ar-C), 128.5 (Ar-C), 128.4 (Ar-C), 128.2 (Ar-C), 127.8 (Ar-C), 126.5 (Ar-C), 110.0 (C-5), 83.9 (C-1'), 82.2 (d, $^3J_{C,P}$ = 7.4 Hz, C-4'), 73.6 (C-3'), 68.7 (2d, $^2J_{C,P}$ = 5.4 Hz, 2 x OCH_2Ph), 67.0 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 53.5 (αC-Phe), 52.4 (αC-Lys), 51.8 (OCH_3), 50.4 (αC-Met), 47.8 (αC-Ala), 40.2

(εC-Lys), 36.6 (βC-Phe), 36.0 (C-2'), 32.2 (βC-Met), 31.5 (βC-Lys), 29.3 (γC-Met), 29.0 (δC-Lys), 22.6 (γC-Lys), 18.2 (CH_3-Ala), 14.6 (-SCH_3), 12.0 (CH_3-Thy); ^{31}P NMR (121 MHz, DMSO-d_6) δ = -0.9; HRMS for $C_{50}H_{64}N_7O_{15}PS$ $[M-H]^-$calcd.: 1064.3846, found: 1064.3867.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Ala-L-Ala-L-Phe-OMe]-2'-deoxythymidine (2.32): Compound **2.32** was prepared according to the general procedure starting from **2.24** (246.2 mg, 0.367 mmol), pentapeptide **2.10** (281 mg, 0.3882 mmol) and Et_3N (0.22 mL, 1.55 mmol) in dry CH_2Cl_2 (20 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 92:6, v/v) to give conjugate **2.32** (380 mg, 86%) as a colorless foam. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.36 (br s, 1H, N*H*-Thy), 8.29 (d, *J* = 8.6 Hz, 1H, N*H*-Met), 8.23 (d, *J* = 7.5 Hz, 1H, N*H*-Phe), 8.07 (d, *J* = 7.6 Hz, 1H, αN*H*-Lys), 8.02 (s, 1H, C*H*O), 7.96 (d, *J* = 7.3 Hz, 1H, αN*H*-Ala), 7.86 (d, *J* = 7.6 Hz, 1H, αN*H*-Ala), 7.50 (s, 1H, H-6), 7.36-7.32 (m, 11H, Ar-*H* of OBn and εN*H*-Lys), 7.29-7.19 (m, 5H, Ar-*H* of Phe), 6.19 (app t, *J* = 7.3 Hz, 1H, H-1'), 5.07-5.04 (m, 5H, H-3' and 2 x OCH_2Ph), 4.49-4.40 (m, 2H, *αH*-Met and *αH*-Phe), 4.29-4.19 (unresolved m, 5H, 2 x *αH*-Ala, *αH*-Lys H-5' and H-5''), 4.12-4.08 (m, 1H, H-4'), 3.57 (s, 3H, OCH_3), 3.05-2.89 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.43 (t, *J* = 7.9 Hz, 2H, γCH_2-Met), 2.25-2.17 (m, 2H, H-2' and H-2''), 2.03 (s, 3H, SCH_3), 1.96-1.75 (m, 2H, βCH_2-Met), 1.69 (s, 3H, CH_3-Thy), 1.67-1.46 (m, 2H, βCH_2-Lys), 1.43-1.34 (m, 2H, δCH_2-Lys), 1.31-1.22 (m, 2H, γCH_2-Lys), 1.20-1.13 (m, 6H, 2 x CH_3-Ala); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.1 (CO-$Ala_{C\text{-}ter}$), 171.7 (CO-Phe), 171.5 (CO-$Ala_{N\text{-}ter}$), 171.1 (CO-Lys), 170.6 (CO-Met), 163.5 (C-4), 161.0 (CHO), 155.4 (OCONH), 150.4 (C-2), 137.0 (1C of Phe), 135.9 (d, $^3J_{C,P}$ = 6.9 Hz, 1C of OCH_2Ph), 135.4 (C-6), 129.0 (Ar-C), 128.5 (Ar-C), 128.4 (Ar-C), 128.2 (Ar-C), 127.8 (Ar-C), 126.5 (Ar-C), 110.0 (C-5), 83.9 (C-1'), 82.2 (d, $^3J_{C,P}$ = 7.3 Hz, C-4'), 73.6 (C-3'), 68.7 (2d, $^2J_{C,P}$ = 5.4 Hz, 2 x OCH_2Ph), 67.0 (d, $^2J_{C,P}$ = 5.3 Hz, C-5'), 53.5 (αC-Phe), 52.5 (αC-Lys), 51.8 (OCH_3), 50.4 (αC-Met), 47.9 (αC-Ala), 47.7 (αC-Ala), 40.2 (εC-Lys), 36.6 (βC-Phe), 36.0 (C-2'), 32.2 (βC-Met), 31.4 (βC-Lys), 29.3 (γC-Met), 29.0 (δC-Lys), 22.7 (γC-Lys), 18.2 (CH_3-Ala), 17.9 (CH_3-Ala), 14.6 (-SCH_3), 12.0 (CH_3-Thy); ^{31}P NMR (121 MHz, DMSO-d_6) δ = -0.9; HRMS for $C_{53}H_{69}N_8O_{16}PS$ $[M-H]^-$calcd.: 1135.4217, found: 1135.4203.

General procedure for the synthesis of 3'-*O*-(peptide carbamate)-2'-deoxythymidine-5'-monophosphate salts (2.19 and 2.33-2.40): To a stirring solution of 5'-*O*-(dibenzylphosphate)-3'-*O*-(peptide carbamate)-2'deoxy-thymidine (1 eq.) in EtOH or MeOH was added 10% Pd/C Degussa (0.1-1.0 eq w/w) and the mixture was hydrogenated at atmospheric pressure using a balloon filled with H_2 for 24 h. The catalyst was removed by filtration and the filtrate was concentrated under reduced pressure. The resulting crude residue was purified RP-HPLC (50 mmol TEAB in 98% H_2O + 2% ACN and 50 mmol TEAB in 50% H_2O + 50% ACN). The collected eluate was lyophilized and freeze-dried repeatedly until constant mass to afford the desired 3'-*O*-(peptide carbamate)-2'-deoxythymidine-5'-monophosphate triethylammonium salts.

3'-*O*-[*N*-For-Gly-L-β-amino-Ala-(β-carbamate)-L-Phe-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.33): Compound **2.33** was obtained as a white solid (149 mg, 81%), according to the general procedure starting from a stirring solution of **2.25** (180 mg, 0.205mmol), Et_3N (0.057 mL, 0.410 mmol) and 10% Pd/C Degussa (18 mg, 10% w/w) in MeOH (15 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.16 (s, 1H, C*H*O), 7.79 (s, 1H, H-6), 7.31-7.19 (m, 5H, Ar-*H* of Phe), 6.32 (dd, *J* = 8.4, 6.6 Hz, 1H, H-1'), 5.22-5.21 (m, 1H, H-3'), 4.72-4.69 (m, 1H, *αH*-Phe), 4.48-4.45 (m, 1H, *αH*-β-amino-Ala), 4.30-4.29 (m, 1H, H-4'), 4.05-4.03 (m, 2H, H-5' and H-5''), 3.91 (d, *J* = 3.4 Hz, 2H, CH_2-Gly), 3.70 (s, 3H, OCH_3), 3.44-3.24 (m, 2H, βCH_2-β-amino-Ala), 3.22-2.95 (m, 2H, βCH_2-Phe), 2.39-2.36 (m, 2H, H-2' and H-2''), 1.88 (s, 3H, CH_3-Thy); ^{13}C NMR (125 MHz,

D_2O) δ = 172.4 (CO-Phe), 170.3 (CO-β-amino-Ala), 170.1 (CO-Gly), 165.9 (C-4), 164.2 (CHO), 156.9 (OCONH), 151.1 (C-2), 136.6 (C-6), 135.6 (1C of Ph), 128.6 (Ar-C), 128.0 (Ar-C), 126.5 (Ar-C), 111.3 (C-5), 84.1 (C-1'), 83.0 (d, $^3J_{C,P}$ = 8.7 Hz, C-4'), 75.8 (C-3'), 64.3 (d, $^2J_{C,P}$ = 5.5 Hz, C-5'), 53.4 (αC-Phe), 52.7 (αC-β-amino-Ala), 52.3 (OCH_3), 40.6 (βC-β-amino-Ala), 40.3 (αC-Gly), 36.0 (βC-Phe), 35.9 (C-2'), 11.0 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 0.3; HRMS for $C_{27}H_{35}N_6O_{14}P$ $[M-H]^-$calcd.: 697.1876, found: 697.1884.

3'-*O*-[*N*-For-Gly-L-Lys-(ε-carbamate)-L-Phe-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.19): Compound **2.19** was obtained as a white solid (255 mg, 83%), according to the general procedure starting from a stirring solution of **2.18** (300 mg, 0.326 mmol), Et_3N (0.091 mL, 0.651 mmol) and 10% Pd/C Degussa (30 mg, 10% w/w) in MeOH (25 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.17 (s, 1H, C*H*O), 7.85 (s, 1H, H-6), 7.35-7.19 (m, 5H, Ar-H of Phe), 6.32 (dd, *J* = 7.5, 6.1 Hz, 1H, H-1'), 5.26-5.23 (m, 1H, H-3'), 4.73-4.70 (m, 1H, *αH*-Phe), 4.31 (br s, 1H, H-4'), 4.27-4.21 (m, 1H, *αH*-Lys), 4.04-4.00 (m, 2H, H-5' and H-5''), 3.96 (s, 2H, CH_2-Gly), 3.73 (s, 3H, OCH_3), 3.12-3.10 (m, 2H, εCH_2-Lys), 3.08-3.02 (m, 2H, βCH_2-Phe), 2.45-2.40 (m, 2H, H-2' and H-2''), 1.93 (s, 3H, CH_3-Thy), 1.69-1.58 (m, 2H, βCH_2-Lys), 1.51-1.44 (m, 2H, δCH_2-Lys), 1.29-1.26 (m, 2H, γCH_2-Lys merged with Et_3N); ^{13}C NMR (125 MHz, D_2O) δ = 177.2 (CO-Phe), 175.3 (CO-Lys), 172.7 (CO-Gly), 172.5 (C-4), 170.5 (CHO), 164.4 (OCONH), 157.3 (C-2), 137.2 (C-6), 136.6 (1C of Ph), 129.1 (Ar-C), 128.2 (Ar-C), 126.5 (Ar-C), 111.8 (C-5), 84.6 (C-1'), 83.0 (d, $^3J_{C,P}$ = 7.5 Hz, C-4'), 75.8 (C-3'), 63.7 (d, $^2J_{C,P}$ = 4.0 Hz, C-5'), 55.7 (αC-Phe), 53.6 (αC-Lys), 53.5 (OCH_3), 43.3 (αC-Gly), 40.7 (εC-Lys), 37.5 (C-2'), 36.2 (βC-Phe), 30.3 (βC-Lys), 28.0 (δC-Lys), 22.0 (γC-Lys), 12.1 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 3.8; HRMS for $C_{30}H_{41}N_6O_{14}P$ $[M-H]^-$ calcd.: 739.2345, found: 739.2336.

3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Phe-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.34): Compound **2.34** was obtained as a white solid (239 mg, 78%), according to the general procedure starting from a stirring solution of **2.26** (300 mg, 0.301 mmol), Et_3N (0.084 mL, 0.603 mmol) and 10% Pd/C Degussa (240 mg, 80% w/w) in MeOH (25 mL). ^{1}H NMR (600 MHz, D_2O) δ = 11.29 (br s, 1H, N*H*-Thy), 8.43 (d, *J* = 7.1 Hz, 1H, N*H*-Lys), 8.35 (d, *J* = 8.3 Hz, 1H, N*H*-Met), 8.01 (s, 1H, C*H*O), 7.90 (s, 1H, H-6), 7.34 (t, *J* = 5.5 Hz, εN*H*-Lys), 6.23 (dd, *J* = 9.0, 5.6 Hz, 1H, H-1'), 5.14-5.13 (m, 1H, H-3'), 4.49-4.45 (m, 1H, *αH*-Met), 4.21-4.17 (m, 1H, *αH*-Lys), 4.05 (br s, 1H, H-4'), 3.86 (br s, 2H, H-5' and H-5''), 3.62 (s, 3H, OCH_3), 2.99-2.96 (m, 2H, εCH_2-Lys), 2.44 (t, *J* = 7.9 Hz, 2H, γCH_2-Met), 2.30-2.13 (m, 2H, H-2' and H-2''), 2.04 (s, 3H, SCH_3), 1.93-1.77 (m, 5H, βCH_2-Met and CH_3-Thy), 1.72-1.60 (m, 2H, βCH_2-Lys), 1.41-1.37 (m, 2H, δCH_2-Lys), 1.32-1.27 (m, 2H, γCH_2-Lys); ^{13}C NMR (150 MHz, D_2O) δ = 172.8 (CO-Lys), 172.5 (CO-Phe), 172.1 (CO-Met), 165.9 (C-4), 163.4 (CHO), 156.9 (OCONH), 151.1 (C-2), 136.5 (C-6), 135.6 (1C of Phe), 128.5 (Ar-C), 128.0 (Ar-C), 126.5 (Ar-C), 111.3 (C-5), 84.2 (C-1'), 83.1 (d, $^3J_{C,P}$ = 9.0 Hz, C-4'), 75.2 (C-3'), 64.4 (d, $^2J_{C,P}$ = 4.0 Hz, C-5'), 53.4 (αC-Phe), 52.9 (αC-Lys), 52.2 (OCH_3), 50.6 (αC-Met), 39.3 (εC-Lys), 36.1 (C-2'), 35.9 (βC-Phe), 30.0 (βC-Lys), 29.7 (βC-Met), 28.3 (γC-Met), 27.5 (δC-Lys), 21.4 (γC-Lys), 13.4 (-SCH_3), 11.0 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = -0.2; HRMS for $C_{33}H_{47}N_6O_{14}PS$ $[M-H]^-$calcd.: 813.2536, found: 813.2543.

3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Ala-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.35): Compound **2.35** was obtained as a white solid (148 mg, 76%), according to the general procedure starting from a stirring solution of **2.27** (190 mg, 0.206 mmol), Et_3N (0.058 mL, 0.413 mmol) and 10% Pd/C Degussa (95 mg, 50% w/w) in MeOH (25 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.21 (s, 1H, C*H*O), 7.97 (s, 1H, H-6), 6.44 (dd, *J* = 8.6, 6.3 Hz, 1H, H-1'), 5.35-5.34 (m, 1H, H-3'), 4.63-4.61 (m, 1H, *αH*-Met), 4.50-4.45 (q, *J* = 7.2 Hz, 1H, *αH*-Ala), 4.41-

4.38 (m, 2H, H-4' and *αH*-Lys), 4.11 (br s, 2H, H-5' and H-5''), 3.84 (s, 3H, OC*H*$_3$), 3.24 (t, *J* = 6.4 Hz, 2H, εC*H*$_2$-Lys), 2.69-2.64 (m, 2H, γC*H*$_2$-Met), 2.58-2.50 (m, 2H, H-2' and H-2''), 2.19 (s, 3H, SC*H*$_3$), 2.18-2.07 (m, 2H, βC*H*$_2$-Met), 2.02 (s, 3H, C*H*$_3$-Thy), 1.94-1.80 (m, 2H, βC*H*$_2$-Lys), 1.65-1.59 (m, 2H, δC*H*$_2$-Lys), 1.53-1.48 (m, 5H, γC*H*$_2$-Lys and C*H*$_3$-Ala),; ^{13}C NMR (125 MHz, D_2O) δ = 174.3 (CO-Ala), 173.0 (CO-Lys), 172.5 (CO-Met), 166.1 (C-4), 163.5 (CHO), 157.0 (OCONH), 151.3 (C-2), 136.9 (C-6), 111.4 (C-5), 84.2 (C-1'), 83.5 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 75.5 (C-3'), 63.8 (d, $^2J_{C,P}$ = 3.4 Hz, C-5'), 53.0 (αC-Lys), 52.2 (OCH_3), 50.7 (αC-Met), 48.1 (αC-Ala), 39.4(εC-Lys), 36.0 (C-2'), 29.9 (βC-Met and βC-Lys), 28.3 (γC-Met), 27.7 (δC-Lys), 21.4 (γC-Lys), 15.2 (CH_3-Ala), 13.5 (SCH_3-Met), 11.1 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 2.0; HRMS for $C_{27}H_{43}N_6O_{14}PS$ [M-H]$^-$calcd.: 737.2222, found: 737.2228.

3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Lys(ε-NH_2)-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium hydrochloride salt (2.36): Compound **2.36** was prepared according to the general procedure starting from a stirring solution of **2.28** (300 mg, 0.270 mmol) and 20% $Pd(OH)_2$/C (150 mg, 50% w/w) in EtOH:H_2O 10:1 (25 mL). After HPLC purification, the resulting triethylammonium salt was further acidified with HCl to stabilize the free amino group to afford **2.36** (173.5 mg, 58%) as a white solid. ^{1}H NMR (600 MHz, D_2O) δ = 8.11 (s, 1H, C*H*O), 7.78 (s, 1H, H-6), 6.34 (dd, *J* = 8.6, 6.2 Hz, 1H, H-1'), 5.23-5.22 (m, 1H, H-3'), 4.49-4.47 (m, 1H, *αH*-Met), 4.38-4.28 (m, 4H, 2 x *αH*-Lys and H-4'), 4.09 (br s, 1H, H-5' and H-5''), 3.73 (s, 3H, OC*H*$_3$), 3.13-3.10 (m, 2H, εC*H*$_2$-Lys), 2.98-2.94 (m, 2H, εC*H*$_2$-Lys), 2.57-2.50 (m, 2H, γC*H*$_2$-Met), 2.43-2.36 (m, 2H, H-2' and H-2''), 2.07 (s, 3H, SC*H*$_3$), 2.05-1.98 (m, 2H, βC*H*$_2$-Met), 1.89 (s, 3H, C*H*$_3$-Thy), 1.88-1.61 (m, 8H, 2 x βC*H*$_2$-Lys and 2 x δC*H*$_2$-Lys), 1.52-1.36 (m, 4H, 2 x γC*H*$_2$-Lys); ^{13}C NMR (150 MHz, D_2O) δ = 175.0 (CO-Lys$_{N\text{-}Ter}$), 173.7 (CO-Lys$_{C\text{-}Ter}$), 172.9 (CO-Met), 166.4 (C-4), 164.0 (CHO), 157.3 (OCONH), 151.5 (C-2), 136.9 (C-6), 111.7 (C-5), 84.7 (C-1'), 83.5 (d, $^3J_{C,P}$ = 8.6 Hz, C-4'), 75.5 (C-3'), 64.9 (d, $^2J_{C,P}$ = 4.0 Hz, C-5'), 53.5 (αC-Lys$_{N\text{-}Ter}$), 52.6 (OCH_3), 52.2 (αC-Lys$_{C\text{-}Ter}$), 51.3 (αC-Met), 39.7 (εC-Lys), 36.4 (C-2'), 30.4 (βC-Lys$_{N\text{-}Ter}$), 30.2 (βC-Met), 29.6 (βC-Lys$_{N\text{-}Ter}$), 28.7 (γC-Met), 25.9 (2 x δC-Lys), 21.9 (2 x γC-Lys), 13.7 (SCH_3), 11.4 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = -0.2; HRMS for $C_{52}H_{68}N_7O_{16}PS$ [M-H]$^-$calcd.: 794.2801, found: 794.2811.

3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.38): Compound **2.38** was obtained as a white solid (217 mg, 79%), according to the general procedure starting from a stirring solution of **2.30** (240 mg, 0.316 mmol), Et_3N (0.088 mL, 0.632 mmol) and 10% Pd/C Degussa (120 mg, 50% w/w) in MeOH (25 mL). ^{1}H NMR (500 MHz, DMSO-*d*$_6$) δ = 11.29 (br s, 1H, N*H*-Thy), 8.43 (d, *J* = 7.1 Hz, 1H, N*H*-Lys), 8.35 (d, *J* = 8.3 Hz, 1H, N*H*-Met), 8.01 (s, 1H, C*H*O), 7.90 (s, 1H, H-6), 7.34 (t, *J* = 5.5 Hz, εN*H*-Lys), 6.23 (dd, *J* = 9.0, 5.6 Hz, 1H, H-1'), 5.14-5.13 (m, 1H, H-3'), 4.49-4.45 (m, 1H, *αH*-Met), 4.21-4.17 (m, 1H, *αH*-Lys), 4.05 (br s, 1H, H-4'), 3.86 (br s, 2H, H-5' and H-5''), 3.62 (s, 3H, OC*H*$_3$), 2.99-2.96 (m, 2H, εC*H*$_2$-Lys), 2.44 (t, *J* = 7.9 Hz, 2H, γC*H*$_2$-Met), 2.30-2.13 (m, 2H, H-2' and H-2''), 2.04 (s, 3H, SC*H*$_3$), 1.93-1.77 (m, 5H, βC*H*$_2$-Met and C*H*$_3$-Thy), 1.72-1.60 (m, 2H, βC*H*$_2$-Lys), 1.41-1.37 (m, 2H, δC*H*$_2$-Lys), 1.32-1.27 (m, 2H, γC*H*$_2$-Lys); ^{13}C NMR (125 MHz, DMSO-*d*$_6$) δ = 172.5 (CO-Lys), 171.1 (CO-Met), 163.8 (C-4), 161.0 (CHO), 155.4 (OCONH), 150.7 (C-2), 136.2 (C-6), 110.2 (C-5), 83.8 (C-1' and C-4'), 75.3 (C-3'), 64.3 (d, $^2J_{C,P}$ = 5.2 Hz, C-5'), 52.1 (αC-Lys), 51.9 (OCH_3), 50.2 (αC-Met), 39.6 (εC-Lys), 36.7 (C-2'), 32.3 (βC-Met), 30.3 (βC-Lys), 29.2 (γC-Met), 28.9 (δC-Lys), 22.8 (γC-Lys), 14.7 (-SCH_3), 12.1 (CH_3-Thy); ^{31}P NMR (202 MHz, DMSO-*d*$_6$) δ = -0.3; HRMS for $C_{24}H_{38}N_5O_{13}PS$ [M-H]$^-$calcd.: 666.1851, found: 666.1854.

3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Ala-L-Phe-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.39): Compound **2.39** was obtained as a white solid (232

mg, 81%), according to the general procedure starting from a stirring solution of **2.31** (280 mg, 0.263 mmol), Et_3N (0.073 mL, 0.525 mmol) and 10% Pd/C Degussa (224 mg, 80% w/w) in MeOH (30 mL). ^{1}H NMR (500 MHz, DMSO-d_6) δ = 11.28 (br s, 1H, N*H*-Thy), 8.37 (d, *J* = 7.8 Hz, 1H, N*H*-Met), 8.27 (d, *J* = 7.5 Hz, 1H, N*H*-Phe), 8.09 (d, *J* = 7.8 Hz, 1H, αN*H*-Lys), 8.02 (s, 1H, C*H*O), 7.95 (d, *J* = 7.7 Hz, 1H, αN*H*-Ala), 7.89 (s, 2H, H-6), 7.31 (d, *J* = 5.4 Hz, 1H, εN*H*-Lys), 7.28-7.19 (m, 5H, Ar-*H* of Phe), 6.23 (dd, *J* = 8.7, 5.9 Hz, 1H, H-1'), 5.14-5.13 (m, 1H, H-3'), 4.47-4.40 (m, 2H, *αH*-Met and *αH*-Phe), 4.29-4.24 (m, 1H, *αH*-Ala), 4.21-4.17 (m, 1H, *αH*-Lys), 4.05 (br s, 1H, H-4'), 3.89-3.86 (m, 2H, H-5' and H-5''), 3.57 (s, 3H, OCH_3), 3.03-2.89 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.43 (t, *J* = 8.0 Hz, 2H, γCH_2-Met), 2.29-2.11 (m, 2H, H-2' and H-2''), 2.02 (s, 3H, SCH_3), 1.93-1.73 (m, 5H, βCH_2-Met and CH_3-Thy), 1.64-1.45 (m, 2H, βCH_2-Lys), 1.40-1.34 (m, 2H, δCH_2-Lys), 1.30-1.20 (m, 2H, γCH_2-Lys), 1.20-1.16 (m, 3H, CH_3-Ala); ^{13}C NMR (125 MHz, DMSO-d_6) δ = 172.3 (CO-Ala), 171.8 (CO-Phe), 171.1 (CO-Lys), 170.7 (CO-Met), 163.8 (C-4), 161.1 (CHO), 155.4 (OCONH), 150.6 (C-2), 137.1 (1C of Phe), 136.2 (C-6), 129.1 (Ar-C), 128.3 (Ar-C), 126.6 (Ar-C), 110.2 (C-5), 83.7 (C-1' and C-4'), 75.2 (C-3'), 64.3 (d, $^2J_{C,P}$ = 5.0 Hz, C-5'), 53.6 (αC-Phe), 52.5 (αC-Lys), 51.9 (OCH_3), 50.5 (αC-Met), 47.9 (αC-Ala), 40.2 (εC-Lys), 36.7 (C-2'), 36.6 (βC-Phe), 32.1 (βC-Met), 31.5 (βC-Lys), 29.4 (γC-Met), 29.1 (δC-Lys), 22.7 (γC-Lys), 18.2 (CH_3-Ala), 14.6 (-SCH_3), 12.1 (CH_3-Thy); ^{31}P NMR (202 MHz, DMSO-d_6) δ = -0.3; HRMS for $C_{36}H_{52}N_7O_{15}PS$ $[M-H]^-$calcd.: 884.2907, found: 884.2920.

3'-*O*-[*N*-For-L-Met-L-Lys(ε-carbamate)-L-Ala-L-Ala-L-Phe-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.40): Compound **2.40** was obtained as a white solid (254 mg, 78%), according to the general procedure starting from a stirring solution of **2.32** (320 mg, 0.281 mmol), Et_3N (0.078 mL, 0.563 mmol) and 10% Pd/C Degussa (256 mg, 80% w/w) in MeOH (30 mL). ^{1}H NMR (500 MHz, DMSO-d_6) δ = 11.28 (br s, 1H, N*H*-Thy), 8.37 (d, *J* = 8.2 Hz, 1H, N*H*-Met), 8.28 (d, *J* = 7.5 Hz, 1H, N*H*-Phe), 8.10 (d, *J* = 7.7 Hz, 1H, αN*H*-Lys), 8.02 (s, 1H, C*H*O), 7.99 (d, *J* = 7.4 Hz, 1H, αN*H*-Ala), 7.89-7.86 (m, 2H, H-6 and αN*H*-Ala), 7.31 (d, *J* = 5.5 Hz, 1H, εN*H*-Lys), 7.28-7.19 (m, 5H, Ar-*H* of Phe), 6.23 (dd, *J* = 8.8, 5.8 Hz, 1H, H-1'), 5.15-5.14 (m, 1H, H-3'), 4.47-4.41 (m, 2H, *αH*-Met and *αH*-Phe), 4.30-4.25 (m, 1H, *αH*-Ala$_{C\text{-}ter}$), 4.25-4.23 (m, 1H, *αH*-Ala$_{N\text{-}ter}$), 4.22-4.17 (m, 1H, *αH*-Lys), 4.05 (br s, 1H, H-4'), 3.89-3.86 (m, 2H, H-5' and H-5''), 3.57 (s, 3H, OCH_3), 3.03-2.92 (m, 4H, εCH_2-Lys and βCH_2-Phe), 2.43 (t, *J* = 8.2 Hz, 2H, γCH_2-Met), 2.30-2.12 (m, 2H, H-2' and H-2''), 2.03 (s, 3H, SCH_3), 1.94-1.74 (m, 5H, βCH_2-Met and CH_3-Thy), 1.68-1.47 (m, 2H, βCH_2-Lys), 1.40-1.35 (m, 2H, δCH_2-Lys), 1.30-1.21 (m, 2H, γCH_2-Lys), 1.20-1.16 (m, 6H, 2 x CH_3-Ala); ^{13}C NMR (125 MHz, DMSO-d_6) δ = 172.2 (CO-Ala$_{C\text{-}ter}$), 171.7 (CO-Phe), 171.5 (CO-Ala$_{N\text{-}ter}$), 171.2 (CO-Lys), 170.7 (CO-Met), 163.8 (C-4), 161.0 (CHO), 155.3 (OCONH), 150.4 (C-2), 137.0 (1C of Phe), 136.1 (C-6), 129.1 (Ar-C), 128.2 (Ar-C), 126.5 (Ar-C), 110.1 (C-5), 83.7 (C-1' and C-4'), 75.2 (C-3'), 64.2 (d, $^2J_{C,P}$ = 4.6 Hz, C-5'), 53.6 (αC-Phe), 52.6 (αC-Lys), 51.8 (OCH_3), 50.4 (αC-Met), 48.0 (αC-Ala$_{C\text{-}ter}$), 47.8 (αC-Ala$_{N\text{-}ter}$), 40.2 (εC-Lys), 36.7 (C-2'), 36.6 (βC-Phe), 32.1 (βC-Met), 31.4 (βC-Lys), 29.3 (γC-Met), 29.1 (δC-Lys), 22.7 (γC-Lys), 18.2 (CH_3-Ala), 17.9 (CH_3-Ala), 14.6 (-SCH_3), 12.0 (CH_3-Thy); ^{31}P NMR (202 MHz, DMSO-d_6) δ = -0.3; HRMS for $C_{39}H_{57}N_8O_{16}PS$ $[M-H]^-$calcd.: 955.3278, found: 955.3293.

3'-*O*-[H-L-Ala-L-Lys(ε-carbamate)-L-Ala-OMe]-2'-deoxythymidine-5'-monophosphate TFA salt (2.41): According to the general procedure, a stirring solution of **2.29** (400 mg, 0.476 mmol) and 10% Pd/C Degussa (40 mg, 10% w/w) in MeOH (15 mL) was hydrogenated to give crude 3'-*O*-[*N*-Boc-L-Ala-L-Lys-(ε-carbamate)-L-Ala-OMe]-2'-deoxythymidine-5'-monophosphate **2.37** (357 mg, quantitative) as a white solid, which was used in the next step without further purification. To a stirring solution of **2.37** (357 mg, 0.476 mmol) in H_2O (3 mL) was added thioanisole (0.056 mL, 0.476 mmol) followed by trifluoroacetic acid (1 mL) at 0 °C. The reaction mixture was stirred at r.t.

for 1.5 h and then the volatiles were removed *in vacuo*. The crude residue was coevaporated 3 times with toluene and purified by RP-Prep HPLC (0.05% TFA in 98% H_2O + 2% ACN and 0.05% TFA in 98% ACN + 2% H_2O). The collected eluate was lyophilized and freeze-dried repeatedly until constant mass to afford **2.41** (258 mg, 71% over two steps) as a white solid. ^{1}H NMR (500 MHz, D_2O) δ = 7.75 (s, 1H, H-6), 6.31 (dd, *J* = 8.7, 6.0 Hz, 1H, H-1'), 5.21-5.20 (m, 1H, H-3'), 4.35-4.30 (m, 1H, *αH*-$Ala_{C\text{-}Ter}$), 4.29 (br s, 1H, H-4'), 4.25-4.22 (m, 1H, *αH*-Lys), 4.09-4.05 (m, 2H, H-5' and H-5''), 4.05-4.02 (m, 1H, *αH*-$Ala_{N\text{-}Ter}$), 3.70 (s, 3H, OC*H*$_3$), 3.12-3.10 (m, 2H, εC*H*$_2$-Lys), 2.40-2.34 (m, 2H, H-2' and H-2''), 1.87 (s, 3H, C*H*$_3$-Thy), 1.76-1.70 (m, 1H, βC*H*$_2$-Lys), 1.52-1.44 (m, 5H, δC*H*$_2$-Lys and C*H*$_3$-$Ala_{N\text{-}Ter}$), 1.43-1.33 (m, 5H, γC*H*$_2$-Lys and C*H*$_3$-$Ala_{C\text{-}Ter}$); ^{13}C NMR (125 MHz, D_2O) δ = 174.2 (CO-$Ala_{C\text{-}Ter}$), 173.1 (CO-Lys), 170.1 (CO-$Ala_{N\text{-}Ter}$), 165.9 (C-4), 157.0 (OCONH), 151.2 (C-2), 136.5 (C-6), 111.3 (C-5), 84.3 (C-1'), 83.1 (d, $^3J_{C,P}$ = 9.0 Hz, C-4'), 75.2 (C-3'), 64.5 (d, $^2J_{C,P}$ = 2.8 Hz, C-5'), 53.1 (αC-Lys), 52.2 (OCH_3), 48.2 (αC-$Ala_{C\text{-}Ter}$), 48.1 (αC-$Ala_{N\text{-}Ter}$), 39.3 (εC-Lys), 36.0 (C-2'), 29.9 (βC-Lys), 27.8 (δC-Lys), 21.3 (γC-Lys), 15.9 (CH_3-$Ala_{N\text{-}Ter}$), 15.1 (CH_3-$Ala_{C\text{-}Ter}$), 11.0 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = -0.3; HRMS for $C_{24}H_{39}N_6O_{13}P$ $[M\text{-}H]^-$calcd.: 649.2240, found: 649.2241.

3'-*O*-[H-L-Ala-L-Lys(ε-carbamate)-L-Ala-OH]-2'-deoxythymidine-5'-monophosphate TFA salt (2.42): To a stirring solution of **2.41** (150 mg, 0.196 mmol) in THF:H_2O:MeOH 1:1:1 (3 mL) was added LiOH (16.2 mg, 0.687 mmol) and the solution was stirred at r.t. for 3 h. The reaction mixture was neutralized with trifluoroacetic acid and the volatiles were removed under reduced pressure. The crude residue was purified by RP-Prep HPLC (0.05% TFA in 98% H_2O + 2% ACN and 0.05% TFA in 98% ACN + 2% H_2O) to afford **2.42** (109 mg, 74%) as a white solid. ^{1}H NMR (600 MHz, D_2O) δ = 7.77 (s, 1H, H-6), 6.33 (dd, *J* = 8.9, 5.9 Hz, 1H, H-1'), 5.22-5.21 (m, 1H, H-3'), 4.32-4.29 (m, 2H, *αH*-$Ala_{N\text{-}Ter}$ and H-4'), 4.27-4.25 (m, 1H, *αH*-Lys), 4.09-4.08 (m, 2H, H-5' and H-5''), 4.07-4.05 (m, 1H, *αH*-$Ala_{C\text{-}Ter}$), 3.12-3.10 (m, 2H, εC*H*$_2$-Lys), 2.42-2.36 (m, 2H, H-2' and H-2''), 1.88 (s, 3H, C*H*$_3$-Thy), 1.81-1.69 (m, 1H, βC*H*$_2$-Lys), 1.53-1.46 (m, 5H, δC*H*$_2$-Lys and C*H*$_3$-$Ala_{C\text{-}Ter}$), 1.44-1.36 (m, 5H, γC*H*$_2$-Lys and C*H*$_3$-$Ala_{N\text{-}Ter}$); ^{13}C NMR (150 MHz, D_2O) δ = 176.0 (CO-$Ala_{C\text{-}Ter}$), 173.3 (CO-Lys), 170.4 (CO-$Ala_{N\text{-}Ter}$), 166.3 (C-4), 157.4 (OCONH), 151.5 (C-2), 136.9 (C-6), 111.7 (C-5), 84.7 (C-1'), 83.5 (d, $^3J_{C,P}$ = 9.0 Hz, C-4'), 75.5 (C-3'), 64.9 (d, $^2J_{C,P}$ = 3.4 Hz, C-5'), 53.4 (αC-Lys), 48.5 (αC-$Ala_{C\text{-}Ter}$ and αC-$Ala_{N\text{-}Ter}$), 39.7 (εC-Lys), 36.4 (C-2'), 30.3 (βC-Lys), 28.1 (δC-Lys), 21.7 (γC-Lys), 16.3 (CH_3-$Ala_{C\text{-}Ter}$), 15.7 (*CH*$_3$-$Ala_{N\text{-}Ter}$), 11.4 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = -0.3; HRMS for $C_{23}H_{37}N_6O_{13}P$ $[M\text{-}H]^-$calcd.: 635.2083, found: 635.2093.

5'-*O*-(4-Monomethoxytrityl)-xylo-2'-deoxythymidine (2.43): To a stirring solution of 2'-deoxythymidine (2.50 g, 10.3 mmol) in pyridine (45 mL) were added Et_3N (1.72 mL, 12.4 mmol), DMAP (0.13 g, 1.03 mmol) followed by 4-monomethoxytrityl chloride (3.66 g, 11.9 mmol) at r.t. and resulting mixture was stirred overnight at r.t. The reaction mixture was cooled and Et_3N (1.72 mL, 12.4 mmol) was added follwed by MsCl (0.88 mL, 11.4 mmol). After stirring for 2 h at r.t., the mixture was filtered, the solid was washed with ethyl acetate and the filtrate was concentrated *in vacuo*. The residue was dissolved in EtOH (45 mL) and then 1M NaOH (25 mL) was added. The mixture was refluxed for 1.5 h, then cooled to r.t. and neutralized with 1N HCl (10 mL). Ethanol was removed *in vacuo* and the residue was extracted with CH_2Cl_2 (3 x 70 mL). The collected organic fractions were washed with brine, dried over Na_2SO_4, concentrated under reduced pressure to leave a crude residue which was purified by column chromatography on silica gel (gradient Hexane/Ethyl acetate 2:1, v/v; 1:1, v/v; 1:2, v/v) to give **2.43** (2.54 g, 48%) as a pale yellow foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.85 (br s,1H, N*H*), 7.61 (s, 1H, H-6), 7.48-7.46 (m, 4H, Ar-H), 7.37-7.20 (m, 8H, Ar-H), 6.86-6.83 (m, 2H, Ar-H), 6.16 (dd, *J* = 7.8, 1.7 Hz, 1H, H-1'), 4.42-4.40 (m, 1H, H-3'), 4.06-4.01 (m, 1H, H-4'), 3.78 (s, 3H, OC*H*$_3$), 3.66-3.61 (m, 1H, H-5'), 3.61-3.46 (m, 1H, H-5''), 2.56-

2.479 (m, 1H, H-2'), 2.29-2.24 (m, 1H, H-2''), 1.72 (s, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 164.4, 158.8, 151.0, 144.1, 144.0, 137.4, 135.1, 130.4, 128.4, 128.3, 128.1, 127.2, 113.4, 109.8, 87.2, 85.6, 83.5, 62.2, 55.3, 41.4, 12.5; HRMS for $C_{30}H_{30}N_2O_6$ $[M+Na]^+$calcd.: 537.1996, found: 537.1995.

3'-*O*-phthalimido-2'-deoxythymidine (2.44): To a stirring suspension of **2.43** (1.85 g, 3.59 mmol), *N*-hydroxyphthalimide (0.79 g, 4.85 mmol) and PPh_3 (1.25 g, 4.75 mmol) in toluene (35 mL), was added DIAD (0.98 g, 4.85 mmol) dropwise at 0°C. The reaction mixture was slowly warmed to r.t. and stirred for 2 h. After completion, the reaction mixture was concentrated *in vacuo* and the residue was dissolved in 80% AcOH (50 mL). After stirring for 3 h at r.t., the reaction mixture was concentrated *in vacuo,* coevaporating with toluene 3 times. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 97:3, v/v) to give **2.44** (0.42 g, 28% over two steps) as a white solid. ^{1}H NMR (300 MHz, DMSO-*d*$_6$) δ = 11.34 (br s,1H, N*H*), 7.89 (s, 4H, Ar-*H*), 7.72 (s, 1H, H-6), 6.39 (dd, *J* = 8.9, 5.6 Hz, 1H, H-1'), 5.20-5.14 (m, 1H, H-3'), 4.97-4.95 (m, 1H, O*H*), 4.23-4.22 (m, 1H, H-4'), 3.62-3.59 (m, 2H, H-5' and H-5''), 2.46-2.26 (m, 1H, H-2' and H-2''), 1.77 (s, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, DMSO-*d*$_6$) δ = 163.7 (C-4), 163.6 (CO-Phthalimide), 150.5 (C-2), 135.9 (C-6), 134.9 (Ar-C), 128.6 (Ar-C), 123.4 (Ar-C), 109.7 (C-5), 88.1 (C-1'), 83.7 (app. C-3'), 82.9 (app. C-4'), 61.4 (C-5'), 35.4 (C-2'), 12.3 (CH_3); HRMS for $C_{18}H_{17}N_3O_7$ $[M+Na]^+$calcd. 410.0959:, found: 410.0958.

5'-*O*-(Dibenzylphosphate)-3'-*O*-phthalimido-2'-deoxythymidine (2.45): Following a similar procedure as used for the synthesis of **2.22**, compound **2.45** was obtained starting from **2.44** (0.40 g, 1.03 mmol), 0.45M tetrazole in ACN (11.47 mL, 5.16 mmol), dibenzyldiisopropyl phosphoramidite (0.78 mL, 2.27 mmol) in dry CH_2Cl_2 (5 mL) and 30% H_2O_2 (0.44 mL, 5.16 mmol). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 100:0, v/v; 99:1, v/v; 98:2, v/v) to give **2.45** (0.62 g, 98%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 10.08 (br s,1H, N*H*), 7.82-7.73 (m, 4H, Ar-*H*), 7.37 (s, 1H, H-6), 7.31-7.30 (m, 10H, Ar-*H*), 6.49 (dd, *J* = 8.2, 6.0 Hz, 1H, H-1'), 5.12-5.00 (m, 4H, OC*H*$_2$Ph), 4.86-4.84 (m, 1H, H-3'), 4.53-4.52 (m, 1H, H-4'), 4.26-4.25 (m, 2H, H-5' and H-5''), 2.69-2.63 (m, 1H, H-2'), 2.10-2.02 (m, 1H, H-2''), 1.80 (s, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 164.1 (C-4), 163.6 (CO-Phthalimide), 150.3 (C-2), 135.3 (C-6), 135.2 (d, $^3J_{C,P}$ = 6.1 Hz, 1C of OCH_2Ph), 134.7 (Ar-C), 128.6 (Ar-C), 128.5 (Ar-C), 128.4 (Ar-C), 127.9 (Ar-C), 127.8 (Ar-C), 123.6 (Ar-C), 111.2 (C-5), 87.8 (C-1'), 84.9 (C-3'), 81.0 (d, $^3J_{C,P}$ = 8.3 Hz, C-4'), 69.6 (app t, $^2J_{C,P}$ = 5.0 Hz, OCH_2Ph), 67.0 (d, $^2J_{C,P}$ =5.4 Hz, C-5'), 36.4 (C-2'), 12.2 (CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.8; HRMS for $C_{31}H_{30}N_3O_{12}P$ $[M+H]^+$calcd.: 648.1742, found: 648.1728.

5'-*O*-(Dibenzylphosphate)-3'-*O*-amino-2'-deoxythymidine (2.46): To a stirring solution of **2.45** (0.57 g, 0.88 mmol) in EtOH (15 mL) was slowly added 4% methylamine in H_2O (2.73 ml, 3.52 mmol) at 0 °C. The mixture was stirred at r.t. for 0.5 h, filtered and washed with CH_2Cl_2. The filtrate was concentrated *in vacuo* and the crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 97:3, v/v) to give **2.46** (0.37 g, 81%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.91 (br s,1H, N*H*), 7.39 (s, 1H, H-6), 7.34 (s, 10H, Ar-H), 6.28 (dd, *J* = 8.2, 5.8 Hz, 1H, H-1'), 5.46 (br s, 2H, ON*H*$_2$), 5.12-4.99 (m, 4H, OC*H*$_2$Ph), 4.22-4.19 (m, 4H, H-3', H-4', H-5' and H-5''), 2.93-2.92 (m, 1H, H-2'), 2.40-2.34 (m, 1H, H-2''), 1.81 (s, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 164.1 (C-4), 150.6 (C-2), 135.4 (d, $^3J_{C,P}$ = 6.3 Hz, 1C of OCH_2Ph), 135.2 (C-6), 128.8 (Ar-C), 128.6 (Ar-C), 128.0 (Ar-C), 111.3 (C-5), 84.8 (C-1'), 83.5 (C-3'), 81.5 (d, $^3J_{C,P}$ = 8.1 Hz, C-4'), 69.7- 69.6 (2d, $^2J_{C,P}$ =5.2 Hz, 2 x OCH_2Ph), 67.9 (d, $^2J_{C,P}$ =5.6 Hz, C-5'), 36.4 (C-2'), 12.2 (-CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.5; HRMS for $C_{31}H_{30}N_3O_{12}P$ $[M+Na]^+$calcd.: 540.1506, found: 540.1501.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Glu(δ-carboxamide)-OMe]-2'-deoxythymidine (2.47): To a stirring solution of **2.46** (120 mg , 0.232 mmol), dipeptide **2.2a** (85.5 mg, 0.267 mmol) and DMAP (1.10 mg, 0.009 mmol) in a mixture of dry DMF (1.0 ml) and dry CH_2Cl_2 (5 ml), was added DCC (77.0 mg, 0.371 mmol) at 0 ºC. The reaction mixture was slowly warmed to r.t. and stirred for 24 h. After removal of all the volatiles under reduced pressure, the residue was filtered and washed with CH_2Cl_2. The combined organic layers were concentrated under reduced pressure and the crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:2, v/v; 97:4, v/v; 94:7, v/v) to give **2.47** (120 mg, 63%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 10.65 (br s, 1H, N*H*), 10.34 (br s, 1H, N*H*), 8.20 (s, 1H, C*H*O), 7.80 (d, *J* = 7.9 Hz, 1H, N*H*-Met), 7.74 (d, *J* = 7.1 Hz, 1H, N*H*-Glu), 7.43 (s, 1H, H-6), 7.34-7.28 (m, 10H, Ar-*H* of OBn), 6.24 (dd, *J* = 8.2, 5.7 Hz, 1H, H-1'), 5.08-5.00 (m, 4H, 2 x OCH_2Ph), 4.74-4.70 (m, 1H, *αH*-Glu), 4.63-4.62 (m, 1H, H-3'), 4.57-4.53 (m, 1H, *αH*-Met), 4.33 (br s, 1H, H-4'), 4.25-4.17 (m, 2H, H-5' and H-5''), 3.70 (s, 3H, OCH_3), 2.60-2.54 (m, 3H, H-2' and γCH_2-Met), 2.32-2.04 (unresolved m, 7H, γCH_2-Glu, βCH_2-Met and SCH_3), 2.00-1-91 (m, 2H, βCH_2-Glu), 1.85-1.79 (m, 4H, H-2'' and CH_3); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 171.8 (CO_2Me-Glu), 171.7 (CO-Met), 170.5 (δCO-Glu), 164.1 (C-4), 162.4 (CHO), 151.0 (C-2), 135.2 (C-6), 135.1 (d, $^3J_{C,P}$ = 6.0 Hz, 1C of OCH_2Ph), 128.8 (Ar-C), 128.6 (Ar-C), 128.0 (Ar-C), 111.7 (C-5), 84.9 (C-3'), 84.7 (C-1'), 81.0 (d, $^3J_{C,P}$ = 6.7 Hz, C-4'), 69.8-69.6 (2d, $^2J_{C,P}$ = 5.3 Hz, 2 x -OCH_2Ph), 67.6 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 52.6 (OCH_3), 51.4 (αC-Met), 51.0 (αC-Glu), 36.0 (C-2'), 31.4 (βC-Met), 29.6 (γC-Met), 29.0 (γC-Glu), 27.2 (βC-Glu), 15.1 (SCH_3), 12.3 (CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.8; HRMS for $C_{36}H_{46}N_5O_{13}PS$ $[M-H]^-$ calcd.: 818.2477, found: 818.2446.

3'-*O*-[*N*-For-L-Met-L-Glu(δ-carboxamide)-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.48): Compound **2.48** was obtained as a white solid (96 mg, 78%), according to the general procedure used for the synthesis of **2.19** and **2.33-2.40**, starting from a stirring solution of **2.47** (120 mg, 0.146 mmol), Et_3N (0.041 mL, 0.293 mmol) and 10% Pd/C Degussa (60 mg, 50% w/w) in MeOH (15 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.16 (s, 1H, C*H*O), 7.98 (s, 1H, H-6), 6.43 (dd, *J* = 9.1, 5.7 Hz, 1H, H-1'), 4.76-4.75 (m, 1H, H-3'), 4.60-4.57 (m, 1H, *αH*-Glu), 4.47-4.45 (m, 1H, *αH*-Met), 4.38 (br s, 1H, H-4'), 3.97-3.96 (m, 2H, H-5' and H-5''), 3.77 (s, 3H, OCH_3), 2.64-2.59 (m, 2H, γCH_2-Met), 2.53-2.49 (m, 1H, H-2'), 2.43-2.37 (m, 1H, H-2''), 2.32-2.10 (unresolved m, 7H, γCH_2-Glu, βCH_2-Met and SCH_3), 2.06-2.01 (m, 2H, βCH_2-Glu), 1.95 (s, 3H, CH_3); ^{13}C NMR (125 MHz, D_2O) δ = 172.9 (αCO-Glu), 172.7 (CO-Met), 171.3 (δCO-Glu), 166.4 (C-4), 163.7 (CHO), 151.4 (C-2), 137.3 (C-6), 111.4 (C-5), 85.7 (C-3'), 84.3 (C-1'), 82.1 (d, $^3J_{C,P}$ = 8.7 Hz, C-4'), 63.7 (d, $^2J_{C,P}$ = 3.5 Hz, C-5'), 52.5 (OCH_3), 51.5 (αC-Met), 50.8 (αC-Glu), 34.7 (C-2'), 30.0 (βC-Met), 28.4 (γC-Met), 28.1 (γC-Glu), 25.3 (βC-Glu), 13.5 (SCH_3), 11.2 (CH_3); ^{31}P NMR (202 MHz, DMSO-d_6) δ = 3.4; HRMS for $C_{22}H_{34}N_5O_{13}PS$ $[M-H]^-$ calcd.: 638.1538, found: 638.1542.

5'-*O*-(*tert*-Butyldiphenylsilyl)-2'deoxythymidine (2.49):[53] To a stirring solution of 2'-deoxythymidine (2.54 g, 10.49 mmol) in dry DMF (20 mL) was added a solution of imidazole (1.57 g, 23.08 mmol) and TBDPS-Cl (2.86 ml, 11.01 mmol) in dry DMF (15 mL) at r.t. After stirring for 4 h at r.t., the reaction mixture was diluted with water (350 ml) and the aqueous layer was extracted with ethyl acetate (3 x 150 mL). The collected organic fractions were washed with brine, dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 95:5, v/v) to give **2.49** (4.17 g, 83%) as a white solid. HRMS for $C_{26}H_{32}N_2O_5Si$ $[M+H]^+$ calcd.: 481.2153, found: 481.2156.

5'-*O*-(*tert*-Butyldiphenylsilyl)-3'-*O*-(methylthiomethyl)-2'-deoxythymidine (2.50):[41] To a stirring solution of **2.49** (4.08 g, 8.50 mmol) in DMSO (27.5 mL), acetic anhydride (19.4 ml) and

acetic acid (6.2 ml) were added. The reaction mixture was stirred at r.t. for 48 h and then concentrated under reduced pressure. The residue was neutralized with saturated aq. $NaHCO_3$ (300ml) and extracted with ethyl acetate (3 x 180 mL). The collected organic fractions were washed with brine, dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient Hexane/Ethyl acetate 4:1, v/v; 3:1, v/v; 2:1, v/v) to give **2.50** (3.3 g, 72%) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.54 (br s,1H, N*H*), 7.70-7.67 (m, 4H, Ar-*H*), 7.47-7.38 (m, 7H, H-6 and Ar-*H*), 6.35 (dd, *J* = 8.5, 5.7 Hz, 1H, H-1'), 4.66-4.56 (m, 3H, H-3' and OCH_2S), 4.09-4.08 (m, 1H, H-4'), 4.00-3.82 (m, 2H, H-5' and H-5''), 2.49-2.43 (m, 1H, H-2'), 2.16-2.04 (m, 4H, SCH_3 and H-2''), 1.66 (s, 3H, CH_3), 1.10 (s, 9H, CH_3-^{t}Bu); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 164.1 (C-4), 150.6 (C-2), 135.6 (Ar-C), 135.4 (Ar-C), 135.2 (C-6), 130.2 (Ar-C), 130.1 (Ar-C), 128.1 (Ar-C), 128.0 (Ar-C), 111.3 (C-5), 85.0 (C-1'), 84.8 (C-4'), 76.2 (C-3'), 73.7 (OCH_2S), 64.1 (C-5'), 37.9 (C-2'), 27.1 ((CH_3)$_3$), 19.4 (CMe_3), 13.9 (SCH_3), 12.2 (CH_3); HRMS for $C_{28}H_{36}N_2O_5SSi$ [M+Na]$^+$calcd. 563.2006:, found: 563.2009.

5'-*O*-(*tert*-Butyldiphenylsilyl)-3'-*O*-(phthalimidooxymethyl)-2'deoxy-thymidine (2.51): To a stirring solution of **2.50** (1.50 g, 2.77 mmol) in dry CH_2Cl_2 (20 mL) was added 1M solution of sulfuryl chloride in CH_2Cl_2 (3.32 mL, 3.32 mmol) at 0 °C. The reaction mixture was slowly warmed to r.t. over 2 h and then concentrated under reduced pressure to give a 3'-*O*-chloromethyluridine derivative as a sticky mass, which was used without further purification. In a separate round-bottom flask, *N*-hydroxyphthalimide (1.81 g, 11.08 mmol) was suspended in dry CH_2Cl_2 (30 mL) and to this DBU (1.45 mL, 9.69 mmol) was then added. After 10 min, the red coloured solution was added to the crude 3'-*O*-chloromethyluridine derivative and the reaction mixture was kept stirring at r.t. for 24 h. It was then diluted with CH_2Cl_2 (100 mL) and 1M aq. acetic acid (30 mL) was added under vigorous stirring. The aqueous layer was discarded and the organic layer was washed with saturated aq. $NaHCO_3$ (2 x 50 ml) and brine, dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 97:3, v/v) to give **2.51** (1.73 g, 95%) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.35 (br s,1H, N*H*), 7.84-7.69 (m, 8H, Ar-*H*), 7.53 (s, 1H, H-6), 7.45-7.37 (m, 6H, Ar-*H*), 6.34 (dd, *J* = 9.2, 5.1 Hz, 1H, H-1'), 5.16 (s, 2H, OCH_2O), 5.11-5.09 (m, 3H, H-3'), 4.18 (br s, 1H, H-4'), 4.14-4.00 (m, 2H, H-5' and H-5''), 2.56-2.50 (m, 1H, H-2'), 2.26-2.17 (m, 2H, H-2''), 1.59 (s, 3H, CH_3), 1.14 (s, 9H, CH_3-^{t}Bu); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 164.0 (C-4), 163.5 (CO-Phthalimide), 150.6 (C-2), 135.7 (Ar-C), 135.4 (Ar-C), 135.3 (C-6), 134.6 (Ar-C), 133.2 (Ar-C), 132.4 (Ar-C), 130.2 (Ar-C), 130.1 (Ar-C), 129.0 (Ar-C), 128.1 (Ar-C), 128.0 (Ar-C), 123.7 (Ar-C), 111.3 (C-5), 97.5 (OCH_2O), 85.3 (C-1'), 84.7 (C-4'), 78.6 (C-3'), 64.6 (C-5'), 37.9 (C-2'), 27.1 ((CH_3)$_3$), 19.5 (CMe_3), 12.1 (CH_3); HRMS for $C_{35}H_{37}N_3O_8Si$ [M+Na]$^+$calcd.: 678.2242, found: 678.2258.

3'-*O*-(Phthalimidooxymethyl)-2'-deoxythymidine (2.52): To a stirring solution of **2.51** (1.73 g, 2.64 mmol) in THF (25 mL (19.4 ml) was added triethylamine trihydrofluoride (1.72 ml, 10.55 mmol) at r.t. The reaction mixture was stirred at r.t. for 36 h and then concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 95:5, v/v) to give **2.52** (0.93 g, 84%) as a white solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.30 (s, 1H, N*H*), 7.88 (s, 4H, Ar-*H*), 7.78 (d, *J* = 1.0 Hz, 1H, H-6), 6.12 (dd, *J* = 8.9, 5.6 Hz, 1H, H-1'), 5.26-5.19 (m, 3H, OCH_2O and 5'-OH), 4.80-4.78 (m, 1H, H-3'), 4.03-4.02 (m, 1H, H-4'), 3.78-3.63 (m, 2H, H-5' and H-5''), 2.38-2.32 (m, 1H, H-2'), 2.25-2.16 (m, 2H, H-2''), 1.79 (d, *J* = 0.9 Hz, 3H, CH_3); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 163.7 (C-4), 163.2 (CO-Phthalimide), 150.4 (C-2), 135.9 (C-6), 134.9 (Ar-C), 128.5 (Ar-C), 123.4 (Ar-C), 109.5 (C-5), 97.8

(OCH_2O), 84.9 (C-1'), 83.9 (C-4'), 78.7 (C-3'), 61.5 (C-5'), 36.6 (C-2'), 12.3 (CH_3); HRMS for $C_{19}H_{19}N_3O_8$ $[M+Na]^+$calcd.: 440.1064, found: 440.1071.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(phthalimidooxymethyl)-2'-deoxythymidine (2.53): Following a similar procedure as used for the synthesis of **2.22**, compound **2.53** was obtained starting from **2.52** (0.93 g, 2.23 mmol), 0.45M tetrazole in ACN (24.75 mL, 11.14 mmol), dibenzyldiisopropyl phosphoramidite (1.52 mL, 4.46 mmol) in dry CH_2Cl_2 (10 mL), and 30% H_2O_2 (0.96 mL, 11.14 mmol). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 100:0, v/v; 99:1, v/v; 98:2, v/v) to give **2.53** (1.39 g, 92%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.5 (app d, 1H, N*H*), 7.84-7.74 (m, 4H, phthamido Ar-*H*), 7.37 (d, *J* = 1.0 Hz, 1H, H-6), 7.38-7.33 (m, 10H, benzyl Ar-*H*), 6.49 (dd, *J* = 9.1, 5.3 Hz, 1H, H-1'), 5.11-5.04 (m, 6H, OCH_2Ph and OCH_2O), 4.85-4.83 (m, 1H, H-3'), 4.53-4.52 (m, 1H, H-4'), 4.55-4.49 (m, 1H, H-5'), 4.30-4.24 (m, 1H, H-5''), 4.20-4.19 (m, 1H, H-4'), 2.37-2.30 (m, 1H, H-2'), 1.90-1.80 (m, 4H, H-2'' and CH_3); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 163.9 (C-4), 163.5 (CO-Phthalimide), 150.6 (C-2), 135.6 (d, $^3J_{C,P}$ = 6.1 Hz, 1C of OCH_2Ph), 135.2 (C-6), 134.7 (Ar-C), 128.9 (Ar-C), 128.8 (Ar-C), 128.7 (Ar-C), 128.2 (Ar-C), 128.1 (Ar-C), 127.8 (Ar-C), 123.7 (Ar-C), 111.6 (C-5), 97.5 (OCH_2O), 84.7 (C-1'), 83.1 (d, $^3J_{C,P}$ = 8.3 Hz, C-4'), 78.3 (C-3'), 69.8-69.7 (2d, $^2J_{C,P}$ = 5.4 Hz, 2 x OCH_2Ph), 67.5 (d, $^2J_{C,P}$ =5.7 Hz, C-5'), 37.2 (C-2'), 12.4 (-CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.5; HRMS for $C_{31}H_{30}N_3O_{12}P$ $[M+Na]^+$calcd.: 570.1612, found: 570.1615.

5'-*O*-(dibenzylphosphate)-3'-*O*-(aminooxymethyl)-2'-deoxythymidine (2.54): Following a similar procedure as used for the synthesis of **2.46**, compound **2.54** was obtained starting from **2.53** (1.3 g, 1.92 mmol), 4% methyl amine in H_2O (5.96 ml, 3.52 mmol) and EtOH (30 mL). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 96:4, v/v) to give **2.54** (0.86 g, 82%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.75 (br s, 1H, N*H*), 7.34-7.33 (m, 11H, H-6 and Ar-*H*), 6.25 (app t, *J* = 6.6 Hz, 1H, H-1'), 5.64 (br s, 2H, NH_2), 5.12-4.99 (m, 4H, 2 x OCH_2Ph), 4.78-4.72 (m, 2H, OCH_2O), 4.31-4.26 (m, 1H, H-3'), 4.22-4.12 (m, 3H, H-4', H-5' and H-5''), 2.41-2.33 (m, 1H, H-2'), 2.03-1.98 (m, 1H, H-2''), 1.81 (s, 3H, CH_3); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 164.0 (C-4), 150.5 (C-2), 135.4 (d, $^3J_{C,P}$ = 6.3 Hz, 1C of OCH_2Ph), 135.3 (C-6), 128.9 (Ar-C), 128.7 (Ar-C), 128.1 (Ar-C), 111.2 (C-5), 98.3 (OCH_2O), 84.8 (C-1'), 82.9 (d, $^3J_{C,P}$ = 7.9 Hz, C-4'), 76.4 (C-3'), 69.8-69.7 (2d, $^2J_{C,P}$ = 5.5 Hz, 2 x OCH_2Ph), 66.5 (d, $^2J_{C,P}$ = 5.5 Hz, C-5'), 38.2 (C-2'), 12.4 (-CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.4; HRMS for $C_{25}H_{30}N_3O_9P$ $[M+H]^+$calcd.: 548.1792, found: 548.1812.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Glu-(methyloxy-δ-carboxamide)-OMe]-2'-deoxythymidine (2.55): Following a similar procedure as used for the synthesis of **2.47**, compound **2.55** was obtained starting from **2.54** (200 mg , 0.365 mmol), dipeptide **2.2a** (134.6 mg, 0.420 mmol), DCC (120.6 mg, 0.584 mmol) and DMAP (1.8 mg, 0.0146 mmol) in a mixture of dry DMF (2.0 ml) and dry CH_2Cl_2 (10 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 94:6, v/v) to give **2.55** (211.0 mg, 68%) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 10.52 (br s, 1H, N*H*), 10.20 (br s,1H, N*H*), 8.18 (s, 1H, C*H*O), 7.90 (d, *J* = 7.0 Hz, 1H, N*H*-Met), 7.74 (d, *J* = 6.7 Hz, 1H, N*H*-Glu), 7.34-7.33 (m, 11H, H-6 and Ar-*H* of OBn), 6.24 (app t, *J* = 6.5 Hz, 1H, H-1'), 5.13-5.00 (m, 4H, 2 x OCH_2Ph), 4.92-4.86 (m, 2H, OCH_2O), 4.79-4.72 (m, 1H, H-3'), 4.54-4.52 (m, 1H, *αH*-Glu), 4.42 (br s, 1H, *αH*-Met), 4.34-4.30 (m, 1H, H-5'), 4.22-4.16 (m, 2H, H-4' and H-5''), 3.71 (s, 3H, OCH_3), 2.56 (t, *J* = 7.2 Hz, 2H, γCH_2-Met), 2.44-2.42 (m, 1H, H-2'), 2.35-2.15 (m, 3H, H-2'' and γCH_2-Glu), 2.14-1.85 (unresolved m, 7H, βCH_2-Met, SCH_3, βCH_2-Glu), 1.79 (s, 3H, CH_3); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 171.9 (αCO-Glu), 171.8 (CO-Met), 170.0 (δCO-Glu), 164.1 (C-4), 162.1 (CHO), 150.8 (C-2), 135.3 (C-6), 135.2 (1C of

OCH_2Ph), 128.9 (Ar-C), 128.7 (Ar-C), 128.0 (Ar-C), 111.4 (C-5), 98.0 (OCH_2O), 84.8 (C-1'), 82.9 (d, $^3J_{C,P}$ = 6.8 Hz, C-4'), 77.1 (C-3'; merged with $CDCl_3$), 68.9-69.8 (2d, $^2J_{C,P}$ = 5.6 Hz, 2 x OCH_2Ph), 66.7 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 52.5 (OCH_3), 51.8 (αC-Met), 51.1 (αC-Glu), 37.7 (C-2'), 31.7 (βC-Met), 29.8 (γC-Met), 29.1 (γC-Glu), 27.3 (βC-Glu), 15.1 (SCH_3), 12.3 (CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.6; HRMS for $C_{37}H_{48}N_5O_{14}PS$ $[M-H]^-$ calcd.: 848.2583, found: 848.2560.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Glu-(methyloxy-δ-carboxamide)-NH_2]-2'-deoxythymidine (2.56): Following a similar procedure as used for the synthesis of **2.47**, compound **2.56** was obtained starting from **2.54** (200 mg , 0.365 mmol), dipeptide **2.2b** (128.3 mg, 0.420 mmol), DCC (120.6 mg, 0.584 mmol) and DMAP (1.8 mg, 0.0146 mmol) in a mixture of dry DMF (2.0 ml) and dry CH_2Cl_2 (10 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:3, v/v; 92:8, v/v) to give **2.56** (186.0 mg, 61%) as a white solid. ^{1}H NMR (300 MHz, CD_3OD) δ = 8.09 (s, 1H, C*H*O), 7.42 (s, 1H, H-6), 7.31(br s, 10H, Ar-*H* of OBn), 6.17 (app t, *J* = 6.9 Hz, 1H, H-1'), 5.06-5.03 (m, 4H, 2 x OCH_2Ph), 4.86 (s, 2H, OCH_2O), 4.52-4.48 (m, 2H, H-3' and *αH*-Glu), 4.39-4.35 (m, 1H, *αH*-Met), 4.27-4.24 (m, 1H, H-4' and H-5'), 4.19-4.15 (m, 2H, H-5''), 2.58-2.50 (m, 2H, γCH_2-Met), 2.47-2.34 (m, 1H, H-2'), 2.23-2.07 (m, 3H, H-2'' and γCH_2-Glu), 2.04-1.84 (unresolved m, 7H, βCH_2-Met, SCH_3, βCH_2-Glu), 1.71 (s, 3H, CH_3); ^{13}C NMR (75 MHz, CD_3OD) δ = 175.8 (αCO-Glu), 173.4 (CO-Met), 172.0 (δCO-Glu), 166.1 (C-4), 164.0 (CHO), 152.2 (C-2), 137.3 (C-6), 137.0 (d, $^3J_{C,P}$ = 6.2 Hz, 1C of OCH_2Ph), 129.9 (Ar-C), 129.7 (Ar-C), 129.2 (Ar-C), 112.0 (C-5), 99.1 (OCH_2O), 86.3 (C-1'), 84.3 (d, $^3J_{C,P}$ = 7.8 Hz, C-4'), 79.2 (C-3'), 71.1-71.0 (2d, $^2J_{C,P}$ = 5.6 Hz, 2 x OCH_2Ph), 68.7 (d, $^2J_{C,P}$ = 5.6 Hz, C-5'), 53.6 (αC-Met), 52.8 (αC-Glu), 38.4 (C-2'), 32.6 (βC-Met), 30.9 (γC-Met), 29.9 (γC-Glu), 28.7 (βC-Glu), 15.3 (SCH_3), 12.5 (CH_3); ^{31}P NMR (121 MHz, CD_3OD) δ = -1.1; HRMS for $C_{36}H_{47}N_6O_{13}PS$ $[M-H]^-$ calcd.: 833.2586, found: 833.2567.

3'-*O*-[*N*-For-L-Met-L-Glu-(methyloxy-δ-carboxamide)-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.57): Compound **2.57** was obtained as a white solid (154 mg, 75%), according to the general procedure used for the synthesis of **2.19** and **2.33-2.40**, starting from a stirring solution of **2.55** (200 mg, 0.235 mmol), Et_3N (0.065 mL, 0.471 mmol) and 10% Pd/C Degussa (100 mg, 50% w/w) in MeOH (20 mL). ^{1}H NMR (500 MHz, DMSO-d_6) δ = 12.51 (br s, 1H, N*H*), 11.24 (br s,1H, N*H*), 9.03 (d, *J* = 8.0 Hz, 1H, N*H*-Met), 8.94 (d, *J* = 6.4 Hz, 1H, N*H*-Glu), 8.04 (s, 1H, C*H*O), 7.80 (s, 1H, H-6), 6.11 (app t, *J* = 5.8 Hz, 1H, H-1'), 4.88-4.77 (m, 2H, OCH_2O), 4.67-4.64 (m, 1H, H-3'), 4.49-4.45 (m, 1H, *αH*-Glu), 4.17-4.13 (m, 1H, *αH*-Met), 4.06-4.02 (m, 1H, H-5'), 3.99 (br s, 1H, H-4'), 3.92-3.90 (m, 2H, H-5''), 3.60 (s, 3H, OCH_3), 2.46 (t, *J* = 7.9 Hz, 2H, γCH_2-Met), 2.34-2.29 (m, 1H, H-2'), 2.26-2.06 (m, 3H, H-2'' and γCH_2-Glu), 2.04 (s, 3H, SCH_3), 2.03-1.86 (m, 4H, βCH_2-Glu and βCH_2-Met), 1.81 (s, 3H, CH_3); ^{13}C NMR (125 MHz, DMSO-d_6) δ = 172.1 (αCO-Glu), 171.6 (CO-Met), 168.7 (δCO-Glu), 164.0 (C-4), 161.2 (CHO), 150.5 (C-2), 136.2 (C-6), 109.8 (C-5), 96.2 (OCH_2O), 83.6 (C-1'), 83.2 (d, $^3J_{C,P}$ = 5.8 Hz, C-4'), 74.1 (C-3'), 62.5 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 52.1 (αC-Met), 52.0 (OCH_3), 50.7 (αC-Glu), 36.9 (C-2'), 32.4 (βC-Met), 29.5 (γC-Met), 29.4 (γC-Glu), 27.0 (βC-Glu), 14.8 (SCH_3), 12.3 (CH_3); ^{31}P NMR (202 MHz, DMSO-d_6) δ = 0.0; HRMS for $C_{23}H_{36}N_5O_{14}PS$ $[M-H]^-$ calcd.: 668.1644, found: 668.1656.

3'-*O*-[*N*-For-L-Met-L-Glu-(methyloxy-δ-carboxamide)-NH_2]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.58): Compound **2.58** was obtained as a white solid (135 mg, 73%), according to the general procedure used for the synthesis of **2.19** and **2.33-2.40**, starting from a stirring solution of **2.56** (180.0 mg, 0.216 mmol), Et_3N (0.06 mL, 0.431 mmol) and 10% Pd/C Degussa (90.0 mg, 50% w/w) in a mixture of EtOH (18 mL) and H_2O (2 mL). ^{1}H NMR (600 MHz, DMSO-d_6) δ = 12.26 (br s, 1H, N*H*), 11.28 (br s,1H, N*H*), 9.26 (d, *J* = 7.7 Hz, 1H, N*H*-Met), 8.37 (d,

J = 7.7 Hz, 1H, N*H*-Glu), 8.07 (s, 1H, C*H*O), 7.80 (s, 1H, H-6), 7.34 (s, 1H, CONH_2-Glu), 7.04 (s, 1H, CONH_2-Glu), 6.13 (app t, *J* = 6.1 Hz, 1H, H-1'), 4.87-4.79 (m, 2H, OCH_2O), 4.65-4.62 (m, 1H, H-3'), 4.40-4.36 (m, 1H, *αH*-Glu), 4.12-4.09 (m, 1H, *αH*-Met), 4.02-4.00 (m, 2H, H-4' and H-5'), 3.92-3.89 (m, 1H, H-5''), 2.48 (t, *J* = 8.0 Hz, 2H, γCH_2-Met), 2.33-2.29 (m, 1H, H-2'), 2.26-2.22 (m, 1H, H-2''), 2.13-2.00 (unresolved m, 5H, γCH_2-Glu and SCH_3), 1.98-1.82 (unresolved m, 7H, βCH_2-Glu, βCH_2-Met and CH_3); ^{13}C NMR (150 MHz, DMSO-d_6) δ = 173.2 (αCO-Glu), 171.2 (CO-Met), 169.0 (δCO-Glu), 163.9 (C-4), 161.6 (CHO), 150.4 (C-2), 136.1 (C-6), 109.7 (C-5), 96.0 (OCH_2O), 83.6 (C-1'), 83.1 (d, $^3J_{C,P}$ = 6.0 Hz, C-4'), 74.3 (C-3'), 62.6 (d, $^2J_{C,P}$ = 4.1 Hz, C-5'), 52.5 (αC-Met), 51.2 (αC-Glu), 36.7 (C-2'), 31.7 (βC-Met), 29.6 (γC-Met), 29.5 (γC-Glu), 28.2 (βC-Glu), 14.7 (SCH_3), 12.2 (CH_3); ^{31}P NMR (202 MHz, DMSO-d_6) δ = 0.0; HRMS for $C_{22}H_{35}N_6O_{13}PS$ $[M-H]^-$ calcd.: 653.1647, found: 653.1644.

5'-*O*-(*tert*-Butyldiphenylsilyl)-3'-*O*-[*N*-For-Gly-L-Glu(*γ*-methyloxy-ester)-L-Ala-OMe]-2'-deoxythymidine (2.59): Following a similar procedure as used for the synthesis of **2.51**, compound **2.59** was obtained in two steps starting from **2.50** (1.69 g, 3.12 mmol), 1M SO_2Cl_2 in CH_2Cl_2 (3.75 mL, 3.75 mmol), followed by reaction with tripeptide **2.5a** (1.24 g, 3.91 mmol) and DBU (0.61 mL, 4.06 mmol) in dry CH_2Cl_2 (50 mL). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 96:4, v/v) to give **2.59** (1.87 g, 74%) as a colorless foam. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.36 (s, 1H, N*H*-Thy), 8.43 (d, *J* = 6.7 Hz, 1H, N*H*-Ala), 8.21 (t, *J* = 5.5 Hz, 1H, N*H*-Gly), 8.11 (d, *J* = 8.3 Hz, 1H, N*H*-Glu), 8.05 (s, 1H, C*H*O), 7.65-7.41 (m, 11H, H-6 and Ar-*H* of TBDPS), 6.14 (app t, *J* = 7.0 Hz, 1H, H-1'), 5.32 (dd, *J* = 19.9, 6.6 Hz, 2H, OCH_2O), 4.50-4.48 (m, 1H, H-3'), 4.40-4.32 (m, 1H, *αH*-Glu), 4.29-4.20 (m, 1H, *αH*-Ala), 4.04-4.01 (m, 1H, H-4'), 3.93-3.82 (m, 2H, H-5'and H-5''), 3.76 (d, *J* = 5.8 Hz, 2H, αCH_2-Gly), 3.57 (s, 3H, OCH_3), 2.39 (t, *J* = 8.1 Hz, 2H, γCH_2-Glu), 2.34-2.25 (m, 1H, H-2'and H-2''), 1.98-1.74 (m, 2H, βCH_2-Glu), 1.51 (s, 3H, CH_3-Thy), 1.28 (d, *J* = 7.3 Hz, 3H, CH_3-Ala), 1.02 (s, 9H, ^{t}Bu); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.8 (CO-Ala), 172.0 (δCO-Glu), 170.7 (αCO-Glu), 168.2 (CO-Gly), 163.5 (C-4), 161.4 (CHO), 150.3 (C-2), 135.4 (C-6), 135.1 (1C of Ph), 134.9 (1C of Ph), 132.7 (Ar-C), 132.3 (Ar-C), 130.1 (Ar-C), 130.0 (Ar-C), 128.0 (Ar-C), 127.9 (Ar-C), 109.7 (C-5), 87.7 (OCH_2O), 83.9 (C-4'), 83.6 (C-1'), 79.1 (C-3'), 63.8 (C-5'), 51.7 (OCH_3), 51.1 (αC-Glu), 47.6 (αC-Ala), 40.5 (αC-Gly), 36.8 (C-2'), 29.7 (γC-Glu), 27.2 (βC-Glu), 26.6 (^{t}Bu), 18.8 (1C of ^{t}Bu), 16.6 (CH_3-Ala), 11.8 (CH_3-Thy); HRMS for $C_{39}H_{51}N_5O_{12}Si$ $[M-H]^-$ calcd.: 808.3230, found: 808.3233.

3'-*O*-[*N*-For-Gly-L-Glu(γ-methyloxy-ester)-L-Ala-OMe]-2'-deoxythymidine (2.60): Following a similar procedure as used for the synthesis of **2.52**, compound **2.60** was obtained starting from **2.59** (0.60 g, 0.74 mmol), $Et_3N{\cdot}3HF$ (0.483 mL, 2.96 mmol) in THF (10 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 96:4, v/v; 92:8, v/v) to give **2.60** (0.34 g, 80%) as a colorless foam. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.31 (s, 1H, N*H*-Thy), 8.44 (d, *J* = 6.7 Hz, 1H, N*H*-Ala), 8.21 (t, *J* = 5.0 Hz, 1H, N*H*-Gly), 8.10 (d, *J* = 8.0 Hz, 1H, N*H*-Glu), 8.06 (s, 1H, C*H*O), 7.68 (s, 1H, H-6), 6.11 (app t, *J* = 7.0 Hz, 1H, H-1'), 5.31 (dd, *J* = 11.6, 6.7 Hz, 2H, OCH_2O), 4.39-4.35 (m, 2H, H-3' and *αH*-Glu), 4.27-4.23 (m, 1H, *αH*-Ala), 3.93-3.90 (m, 1H, H-4'), 3.77 (d, *J* = 6.8 Hz, 2H, αCH_2-Gly), 3.61-3.58 (m, 5H, H-5', H-5'' and OCH_3), 2.40 (t, *J* = 8.0 Hz, 2H, γCH_2-Glu), 2.25-2.19 (m, 1H, H-2'and H-2''), 2.00-1.80 (m, 2H, βCH_2-Glu), 1.78 (s, 3H, CH_3-Thy), 1.29 (d, *J* = 7.3 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.8 (CO-Ala), 172.0 (δCO-Glu), 170.7 (αCO-Glu), 168.3 (CO-Gly), 163.6 (C-4), 161.4 (CHO), 150.4 (C-2), 135.9 (C-6), 109.5 (C-5), 87.6 (OCH_2O), 84.7 (C-4'), 83.6 (C-1'), 79.3 (C-3'), 61.1 (C-5'), 51.8 (OCH_3), 51.1 (αC-Glu), 47.6 (αC-Ala), 40.5 (αC-Gly), 36.9 (C-2'), 29.8 (γC-Glu), 27.2 (βC-Glu), 16.6 (CH_3-Ala), 12.2 (CH_3-Thy); HRMS for $C_{23}H_{33}N_5O_{12}$ $[M-H]^-$ calcd.: 570.2053, found: 570.2054.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-Gly-L-Glu(*γ*-methyloxy-ester)-L-Ala-OMe]-2'-deoxythymidine (2.61): Following a similar procedure as used for the synthesis of **2.22**, compound **2.61** (383 mg, 86%) was obtained starting from **2.60** (306 mg, 0.535 mmol) as a colorless foam. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.33 (s, 1H, N*H*-Thy), 8.43 (d, *J* = 6.7 Hz, 1H, N*H*-Ala), 8.20 (t, *J* = 5.7 Hz, 1H, N*H*-Gly), 8.10 (d, *J* = 8.1 Hz, 1H, N*H*-Glu), 8.06 (s, 1H, C*H*O), 7.46 (d, *J* = 0.9 Hz, 1H, H-6), 7.36-7.35 (m, 10H, Ar-*H* of OBn), 6.12 (app t, *J* = 7.1 Hz, 1H, H-1'), 5.29 (dd, *J* = 15.7, 6.6 Hz, 2H, OCH_2O), 5.05 (d, *J* = 8.1 Hz, 4H, 2 x OCH_2Ph), 4.37-4.29 (m, 2H, H-3' and *αH*-Glu), 4.27-4.19 (m, 3H, *αH*-Ala, H-5'and H-5''), 4.10-4.07 (m, 1H, H-4'), 3.77 (d, *J* = 5.8 Hz, 2H, αCH_2-Gly), 3.58 (s, 3H, OCH_3), 2.41 (t, *J* = 7.8 Hz, 2H, γCH_2-Glu), 2.26-2.13 (m, 1H, H-2'and H-2''), 1.99-1.74 (m, 2H, βCH_2-Glu), 1.51 (d, *J* = 0.9 Hz, 3H, CH_3-Thy), 1.28 (d, *J* = 7.3 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.8 (CO-Ala), 171.9 (δCO-Glu), 170.7 (αCO-Glu), 168.3 (CO-Gly), 163.6 (C-4), 161.4 (CHO), 150.3 (C-2), 135.9 (d, $^3J_{C,P}$ = 6.8 Hz, 1C of OCH_2Ph), 135.7 (C-6), 128.5 (Ar-C), 128.4 (Ar-C), 127.8 (Ar-C), 109.9 (C-5), 87.8 (OCH_2O), 84.0 (C-1'), 82.0 (d, $^3J_{C,P}$ = 7.4 Hz, C-4'), 78.8 (C-3'), 68.7 (d, $^2J_{C,P}$ = 5.4 Hz, 2 x OCH_2Ph), 66.8 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 51.8 (OCH_3), 51.1 (αC-Glu), 47.6 (αC-Ala), 40.5 (αC-Gly), 36.5 (C-2'), 29.7 (γC-Glu), 27.2 (βC-Glu), 16.6 (CH_3-Ala), 12.0 (CH_3-Thy); ^{31}P NMR (121 MHz, DMSO-d_6) δ = -0.9; HRMS for $C_{37}H_{46}N_5O_{15}P$ $[M+H]^+$ calcd.: 832.2800, found: 832.2803.

3'-*O*-[*N*-For-Gly-L-Glu(*γ*-methyloxy-ester)-L-Ala-OMe]-2'-deoxythymidine5'-monophosphate triethylammonium salt (2.62): Compound **2.62** was obtained as a white solid (63.6 mg, 62%), according to the general procedure used for the synthesis of **2.19** and **2.33-2.40**, starting from a stirring solution of **2.61** (100 mg, 0.120 mmol), $NaHCO_3$ (20.2 mg, 0.240 mmol) and 20% $Pd(OH)_2$/C (10.0 mg, 10% w/w) in EtOH/H_2O 9:1 (10 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.14 (s, 1H, C*H*O), 7.81 (s, 1H, H-6), 6.26 (app t, *J* = 7.2 Hz, 1H, H-1'), 5.38 (dd, *J* = 44.5, 6.9 Hz, 2H, OCH_2O), 4.62-4.61 (m, 1H, H-3'), 4.39-4.34 (m, 2H, *αH*-Glu and *αH*-Ala), 4.26-4.25 (m, 1H, H-4'), 3.98-3.94 (m, 4H, αCH_2-Gly and H-5'and H-5''), 3.69 (s, 3H, OCH_3), 2.56 (t, *J* = 7.4 Hz, 2H, γCH_2-Glu), 2.46-2.34 (m, 1H, H-2'and H-2''), 2.17-1.95 (m, 2H, βCH_2-Glu), 1.88 (d, *J* = 0.9 Hz, 3H, CH_3-Thy), 1.38 (d, *J* = 7.3 Hz, 3H, CH_3-Ala); ^{13}C NMR (125 MHz, D_2O) δ = 174.5 (CO-Ala), 174.0 (δCO-Glu), 172.8 (αCO-Glu), 170.7 (CO-Gly), 166.4 (C-4), 164.6 (CHO), 151.5 (C-2), 137.4 (C-6), 111.5 (C-5), 88.3 (OCH_2O), 84.6 (C-1'), 83.9 (d, $^3J_{C,P}$ = 9.0 Hz, C-4'), 80.3 (C-3'), 64.0 (d, $^2J_{C,P}$ = 4.4 Hz, C-5'), 52.6 (OCH_3), 52.3 (αC-Glu), 48.5 (αC-Ala), 40.7 (αC-Gly), 36.8 (C-2'), 29.5 (γC-Glu), 25.6 (βC-Glu), 15.5 (CH_3-Ala), 11.4 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 1.8; HRMS for $C_{23}H_{34}N_5O_{15}P$ $[M+H]^+$ calcd.: 650.1716, found: 650.1718.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-L-Met-L-Lys(Boc)-L-Lys(Boc)-ester]-2'-deoxythymidine (2.63): Following a similar procedure as used for the synthesis of **2.47**, compound **2.63** was obtained starting from **2.23** (250 mg, 0.498 mmol), tripeptide **2.4h** (378.4 mg, 0.597 mmol), DCC (164.3 mg, 0.796 mmol) and DMAP (6.1 mg, 0.050 mmol) in a mixture of dry DMF (3 ml) and dry CH_2Cl_2 (15 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 97:2, v/v; 94:4, v/v) to give **2.63** (423 mg, 76%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.51 (br s,1H, N*H*), 9.31 (br s, 1H, N*H*), 8.22 (s, 1H, C*H*O), 7.41 (s, 1H, H-6), 7.35-7.34 (m, 10H, Ar-*H* of OBn), 6.33 (dd, *J* = 9.2, 5.4 Hz, 1H, H-1'), 5.18-5.12 (m, 1H, H-3'), 5.10-4.95 (m, 4H, OCH_2Ph), 4.89-4.77 (m, 1H, *αH*-Met), 4.73-4.34 (m, 2H, 2 x *αH*-Lys), 4.22-4.10 (m, 3H, H-4', H-5' and H-5''), 3.15-3.04 (m, 4H, 2 x εCH_2-Lys), 2.56 (t, *J* = 6.7 Hz, 2H, γCH_2-Met), 2.35-2.22 (m, 1H, H-2'), 2.13-2.07 (m, 4H, H-2'' and SCH_3), 1.96-1.87 (m, 2H, βCH_2-Met), 1.82 (s, 3H, CH_3-Thy), 1.77-1.64 (m, 4H, 2 x βCH_2-Lys), 1.54-1.31 (m, 26H, 2 x δCH_2-Lys, 2 x γCH_2-Lys and 2 x CH_3-^{t}Bu); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 171.9 (CO-$Lys_{N\text{-}Ter}$), 171.7 (CO-Met), 171.3 (CO-$Lys_{C\text{-}Ter}$), 163.7 (C-4), 161.9 (CHO), 156.4 (2 x OCONH), 150.9 (C-2), 135.5 (d, $^3J_{C,P}$ =

6.0 Hz, 1C of OCH_2Ph), 135.1 (C-6), 129.1 (Ar-C), 128.9 (Ar-C), 128.2 (Ar-C), 112.1 (C-5), 84.5 (C-1'), 82.6 (d, $^3J_{C,P}$ = 8.9 Hz, C-4'), 79.4 (2 x 1C of ^{t}Bu), 75.7 (C-3'), 70.0 (app t, $^2J_{C,P}$ = 5.6 Hz, 2 x OCH_2Ph), 67.3 (d, $^2J_{C,P}$ = 5.8 Hz, C-5'), 53.5 (2 x αC-Lys), 52.8 (αC-Met), 40.1 (2 x εC-Lys), 37.0 (C-2'), 31.8 (βC-Met), 31.0 (2 x βC-Lys), 30.1 (γC-Met), 29.5 (2 x δC-Lys), 28.6 (2 x CH_3-^{t}Bu), 22.9 (2 x γC-Lys), 15.4 (SCH_3), 12.5 (CH_3-Thy); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.7; HRMS for $C_{52}H_{76}N_7O_{16}PS$ $[M+H]^+$calcd.: 1118.4879, found: 1118.4878.

3'-*O*-[*N*-For-L-Met-L-Lys-L-Lys-ester]-2'-deoxythymidine-5'-monophosphate TFA salt (2.65): Following a similar procedure as used for the synthesis of **2.41**, compound **2.65** was obtained as a white solid in two steps (214 mg, 62%) starting from **2.63** (400 mg, 0.358 mmol) and 10% Pd/C Degussa (200 mg, 50% w/w) in EtOH:H_2O 10:1 (30 mL), to give crude 3'-*O*-[*N*-For-Met-L-Lys(Boc)-L-Lys(Boc)-ester]-2'-deoxythymidine-5'-monophosphate **2.64** (335.5 mg, quantitative) as a white solid, which was immediately deprotected in H_2O (4.5 mL), thioanisole (0.056 mL, 0.476 mmol) and trifluoroacetic acid (1.5 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.11 (s, 1H, C*H*O), 7.80 (s, 1H, H-6), 6.38-6.33 (m, 1H, H-1'), 5.47-5.44 (m, 1H, H-3'), 4.52-4.48 (m, 1H, *αH*-Met), 4.46-4.33 (m, 3H, 2 x *αH*-Lys and H-4'), 4.14-4.09 (m, 2H, H-5' and H-5''), 3.00-2.98 (m, 4H, 2 x εCH_2-Lys), 2.61-2.51 (m, 2H, γCH_2-Met), 2.48-2.42 (m, 1H, H-2' and H-2''), 2.09 (m, 3H, SCH_3), 2.05-1.97 (m, 2H, βCH_2-Met), 1.91 (s, 3H, CH_3-Thy), 1.85-1.78 (m, 4H, 2 x βCH_2-Lys), 1.72-1.66 (m, 4H, 2 x δCH_2-Lys), 1.52-1.41 (m, 4H, 2 x γCH_2-Lys); ^{13}C NMR (125 MHz, D_2O) δ = 173.3 ($CO\text{-}Lys_{N\text{-}Ter}$), 172.7 (CO-Met), 171.7 ($CO\text{-}Lys_{C\text{-}Ter}$), 166.0 (C-4), 163.7 (CHO), 151.2 (C-2), 136.5 (C-6), 111.4 (C-5), 84.2 (C-1'), 82.7 (d, $^3J_{C,P}$ = 8.9 Hz, C-4'), 76.1 (C-3'), 64.5 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 52.9 (αC-Lys), 52.0 (αC-Lys), 50.8 (αC-Met), 38.5 (2 x εC-Lys), 35.7 (C-2'), 29.8 (βC-Met), 28.8 (2 x βC-Lys), 28.4 (γC-Met), 25.5 (2 x δC-Lys), 21.5 (2 x γC-Lys), 13.5 (SCH_3), 11.0 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = -0.2; HRMS for $C_{28}H_{48}N_7O_{12}PS$ $[M-H]^-$calcd.: 736.2746, found: 736.2742.

3',5'-*O*-Di(*tert*-Butyldimethylsilyl)-2'deoxy-thymidine (2.66):[51] To a stirring solution of thymidine (2.00 g, 8.26 mmol) in dry DMF was added imidazole (3.37 g, 49.5 mmol) followed by TBDMSCl (3.73 g, 24.8 mmol) and the solution was stirred overnight at r.t. The reaction mixture was then diluted with CH_2Cl_2, washed with water and dried over Na_2SO_4, filtered and evaporated *in vacuo*. The remaining crude residue was purified by column chromatography on silica gel (gradient Hexane/EtOAc 90:10, v/v; 80:20, v/v; 70:30, v/v) to give **2.66** (3.03 g, 78%) as a white foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ =9.23 (br s, 1H, N*H*), 7.45 (s, 1 H, H-6), 6.34-4.30 (m, 1H, H-1'), 4.39-4.37 (m, 1H, H-3'), 3.91-3.83 (m, 2H, H-5' and H-4'), 3.76-3.71 (m, 1H, H-5''), 2.27-2.20 (m, 1H, H-2'), 2.02-1.96 (m, 1H, H-2''), 1.89 (s, 3H, CH_3-Thy), 0.91 (s, 9H, C(CH_3)$_3$), 0.97 (s, 9H, C(CH_3)$_3$), 0.09 (s, 6H, CH_3), 0.06 (s, 6H, CH_3); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 163.6 (C-4), 150.1 (C-2), 135.1 (C-6), 110.5 (C-5), 87.5 (C-4'), 84.5 (C-1'), 71.9 (C-3'), 62.6 (C-5'), 41.5 (C-2'), 25.6 (^{t}Bu), 25.4 (^{t}Bu), 18.1 (1C of ^{t}Bu), 17.7 (1C of ^{t}Bu), 12.2 (CH_3-Thy), -5.0 (CH_3), -5.2 (CH_3), -5.7 (CH_3), -5.8 (CH_3); HRMS for $C_{22}H_{43}N_2O_5Si_2$ $[M+H]^+$calcd.: 471.2705, found: 471.2706.

3'-*O*-(*tert*-Butyldimethylsilyl)-2'deoxy-thymidine (2.67):[51] Compound **2.66** (2.39 g, 5.08 mmol) was dissolved in CH_2Cl_2 (28 mL) and the solution was cooled to 0 °C. A cooled TFA:H_2O (8:1, 2.8 mL) mixture was then added dropwise and the reaction mixture was stirred for 4 h at 0 °C. The solution was then diluted with cooled CH_2Cl_2, washed with ice-cold water and brine, dried over Na_2SO_4, filtered and evaporated *in vacuo*. The remaining crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 95:5, v/v) to give **2.67** (1.07 g, 59%) as a white foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.40 (br s, 1H, N*H*), 7.38 (d, *J* = 1.08 Hz, 1H, H-6), 6.13 (t, *J* = 6.7 Hz, 1H, H-1'), 4.48-4.43 (m, 1H, H-3'), 3.89-3.86 (m, 2H, H-5' and H-5''), 3.73-3.72 (m, 1H, H-4'), 3.07 (br s, 1H, O*H*), 2.29-2.27 (m, 1H, H-2'), 2.22-2.18 (m, 1H, H-2''), 1.85 (s, 3H,

C*H*$_3$-Thy), 0.85 (s, 9H, C(C*H*$_3$)$_3$), 0.05 (s, 6H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 164.1 (C-4), 150.4 (C-2), 137.0 (C-6), 110.8 (C-5), 87.5 (C-4'), 86.5 (C-1'), 71.5 (C-3'), 61.8 (C-5'), 40.5 (C-2'), 25.6 (^{t}Bu), 17.9 (1C of ^{t}Bu), 12.4 (CH_3-Thy), -4.8 (CH_3), -4.9 (CH_3); HRMS for $C_{16}H_{29}N_2O_5Si$ $[M+H]^+$calcd.: 357.1840, found: 357.1841.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(*tert*-butyldimethylsilyl)-2'-deoxythymidine (2.68): Following a similar procedure as used for the synthesis of **2.22**, compound **2.68** was obtained starting from **2.67** (0.47 g, 1.32 mmol), 0.45M tetrazole in ACN (15 mL, 6.57 mmol), dibenzyldiisopropyl phosphoramidite (0.95 mL, 2.87 mmol) in dry CH_2Cl_2 (10 mL), and 35% H_2O_2 (0.57 mL, 7.01 mmol). The crude residue was purified by column chromatography on silica gel (gradient Hexane/EtOAc 90:10, v/v; 70:30, v/v; 50:50, v/v) to give **2.68** (0.75 g, 88%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.11 (br s,1H, N*H*), 7.33-7.32 (m, 11H, Ar-*H* and H-6), 6.29 (dd, *J* = 7.5, 6.1 Hz, 1H, H-1'), 5.11-4.98 (m, 4H, OC*H*$_2$Ph), 4.32-4.28 (m, 1H, H-3'), 4.15-4.09 (m, 2H, H-5' and H-5''), 3.95-3.93 (m, 1H, H-4'), 2.15 (ddd, *J* = 13.4, 6.0, 3.1 Hz, 1H, H-2'), 1.93-1.84 (m, 1H, H-2''), 1.82 (d, *J* = 0.9 Hz, 3H, C*H*$_3$-Thy), 0.86 (s, 9H, C(C*H*$_3$)$_3$), 0.04 (s, 3H, C*H*$_3$), 0.03 (s, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 163.5 (C-4), 150.0 (C-2), 135.2 (1C of OCH_2Ph), 135.0 (C-6), 128.5 (Ar-C), 128.4 (Ar-C), 127.7 (Ar-C), 110.9 (C-5), 85.0 (d, $^3J_{C,P}$ = 8.0 Hz, C-4'), 84.5 (C-1'), 71.5 (C-3'), 69.4 (2 x d, $^2J_{C,P}$ = 5.4 Hz, OCH_2), 66.2 (d, $^2J_{C,P}$ =5.8 Hz, C-5'), 40.4 (C-2'), 25.3 (^{t}Bu), 17.6 (1C of ^{t}Bu), 12.0 (CH_3-Thy), -5.0 (CH_3), -5.2 (CH_3); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.5; HRMS for $C_{30}H_{42}N_2O_8PSi$ $[M+H]^+$calcd.: 617.2442, found: 617.2447.

5'-*O*-(Dibenzylphosphate)-2'-deoxythymidine (2.23): Method (a). To a solution of **2.68** (570 mg, 0.92 mmol) in THF (10 mL) cooled at 0 ºC was added TBAF (1M in THF, 0.92 mL, 0.92 mmol). The reaction mixture was stirred for 45 min at 0 ºC and for 15 min at r.t. The mixture was partially evaporated and then purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 96:4, v/v) to give **2.23** (227 mg, 49%) as a colorless foam. **Method (b)**. To a solution of **2.68** (0.80 g, 1.30 mmol) in THF (10 mL) was added $Et_3N{\cdot}3HF$ (0.53 mL, 3.24 mmol) and the solution was stirred at r.t. for 24 h. After removal of all the volatiles, the resulting residue was redissolved in CH_2Cl_2 and washed with 10% NaHCO3 solution. The organic layer was then dried over Na_2SO_4, filtered and evaporated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 98:2, v/v; 96:4, v/v) to give **2.23** (0.58 g, 89%) as a colorless foam.

5'-*O*-(Dibenzylphosphate)-3'-*O*-[*N*-For-Gly-L-Glu(*δ*-ester)-L-Phe-OMe]-2'-deoxythymidine (2.69): Following a similar procedure as used for the synthesis of **2.47**, compound **2.69** was obtained starting from **23** (230 mg, 0.458 mmol), tripeptide **2.5b** (200.0 mg, 0.508 mmol), DCC (110.0 mg, 0.533 mmol) and DMAP (1.1 mg, 0.090 mmol) in a mixture of dry DMF (4 ml) and dry CH_2Cl_2 (4 ml). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH 99:1, v/v; 95:5, v/v; 93:7, v/v) to give **2.69** (230 mg, 57%) as a white foam. ^{1}H NMR (300 MHz, MeOD) δ = 8.18 (s, 1H, C*H*O), 7.46 (s, 1H, H-6), 7.36 (br s, 10H, Ar-*H* of OBn), 7.29-7.19 (m, 5H, Ar-*H* of Phe), 6.25 (dd, *J* = 8.6, 5.7 Hz, 1H, H-1'), 5.21-5.20 (m, 1H, H-3'), 5.13-5.09 (m, 4H, OC*H*$_2$Ph), 4.68 (dd, *J* = 8.5, 5.7 Hz, 1H, αC*H*-Phe), 4.68 (dd, *J* = 8.3, 5.7 Hz, 1H, αC*H*-Glu), 4.28-4.26 (m, 2H, H-5' and H-5''), 4.19-4.18 (m, 1H, H-4'), 3.93 (br s, 2H, C*H*$_2$-Gly), 3.68 (s, 3H, OC*H*$_3$), 3.16 (dd, *J*= 13.7, 5.6 Hz, 1H, βC*H*$_2$-Phe), 3.00 (dd, *J*= 13.9, 8.6 Hz, 1H, βC*H*$_2$-Phe), 2.43 (t, *J* = 7.1 Hz, 2H, γC*H*$_2$-Glu), 2.37-2.33 (m, 1H, H-2'), 2.15-2.06 (m, 2H, H-2'' and βC*H*$_2$-Glu), 1.96-1.86 (m, 1H, βC*H*$_2$-Glu), 1.77 (s, 3H, C*H*$_3$-Thy); ^{13}C NMR (75 MHz, MeOD) δ = 172.0 (δCO-Glu), 171.5, 171.4, 169.1, 164.4 (C-4), 162.6 (CHO), 150.5 (C-2), 135.4 (C-6), 135.3 (d, $^3J_{C,P}$ = 6.1 Hz, 1C of OCH_2Ph), 128.6 (Ar-C), 128.2 (Ar-C), 128.1 (Ar-C), 127.8 (Ar-C), 127.5 (Ar-C), 126.2 (Ar-C), 110.4

(C-5), 84.5 (C-1'), 82.2 (d, $^{3}J_{C,P}$ = 7.8 Hz, C-4'), 74.1 (C-3'), 69.4 (d, $^{2}J_{C,P}$ = 5.7 Hz, 2 x OCH_2Ph), 66.9 (d, $^{2}J_{C,P}$ = 5.8 Hz, C-5'), 53.5 (αC-Phe), 51.6 (αC-Glu), 51.1 (OCH_3), 40.3 (αC-Gly), 36.5 (βC-Phe), 36.0 (C-2'), 29.0 (γC-Glu), 26.4 (βC-Glu), 10.8 (CH_3-Thy); ^{31}P NMR (121 MHz, MeOD) δ = -1.1; HRMS for $C_{42}H_{49}N_5O_{14}P$ $[M+H]^+$calcd.: 878.3008, found: 878.3006.

3'-*O*-[*N*-For-L-Gly-L-Glu(*δ*-ester)-L-Phe-OMe]-2'-deoxythymidine-5'-monophosphate triethylammonium salt (2.70): Compound **2.70** was obtained as a white solid (160 mg, 66%), according to the general procedure used for the synthesis of **2.19** and **2.33-2.40**, starting from a stirring solution of **2.69** (190 mg, 0.216 mmol) and 10% Pd/C Degussa (120 mg, 50% w/w) in THF (20 mL). ^{1}H NMR (300 MHz, D_2O) δ = 8.15 (s, 1H, C*H*O), 7.70 (s, 1H, H-6), 7.34-7.20 (m, 5H, Ar-*H* of Phe), 6.36 (dd, *J* = 9.1, 5.5 Hz, 1H, H-1'), 5.35 (d, *J* = 5.2 Hz, 1H, H-3'), 4.68 (dd, *J* = 9.5, 5.5 Hz, 1H, αC*H*-Phe), 4.29-4.24 (m, 2H, αC*H*-Glu and H-4'), 4.93-3.92 (m, 4H, αCH_2-Gly, H-5' and H-5''), 3.70 (s, 3H, OCH_3), 3.25-3.18 (m, 1H, βCH_2-Phe), 2.99-2.91 (m, 1H, βCH_2-Phe), 2.42-2.88 (m, 4H, γCH_2-Glu, H-2' and H-2''), 1.98-1.82 (m, 2H, βCH_2-Glu), 1.86 (s, 3H, CH_3-Thy); ^{13}C NMR (151 MHz, D_2O) δ = 173.6, 172.1, 170.4, 169.1, 164.4 (CHO), 163.8 (C-4), 153.6 (C-2), 137.4 (1C of Ph), 137.1 (C-6), 129.0 (Ar-C), 128.3 (Ar-C), 126.5 (Ar-C), 111.5 (C-5), 85.6 (d, $^{3}J_{C,P}$ = 8.0 Hz, C-4'), 84.5 (C-1'), 71.0 (C-3'), 63.4 (d, $^{2}J_{C,P}$ = 4.06 Hz, C-5'), 53.7 (αC-Phe), 55.8 (αC-Glu), 53.5 (OCH_3), 41.6 (αC-Gly), 38.0 (βC-Phe), 37.4 (C-2'), 28.6 (γC-Glu), 26.7 (βC-Glu), 11.7 (CH_3-Thy); ^{31}P NMR (121 MHz, D_2O) δ = -3.7; HRMS for $C_{28}H_{35}N_5O_{14}P$ $[M-H]^-$ calcd.: 696.1923, found: 696.1926.

2.5. REFERENCES

[1] V. B. Pinheiro, P. Holliger, *Curr. Opin. Chem. Biol.* **2012**, *16*, 245-252.
[2] J. C. Chaput, H. Yu, S. Zhang, *Chem. Biol.* **2012**, *19*, 1360-1371.
[3] P. Herdewijn, P. Marliere, *Chem. Biodivers.* **2009**, *6*, 791-808.
[4] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, P. Holliger, *Science* **2012**, *336*, 341-344.
[5] V. Pezo, F. W. Liu, M. Abramov, M. Froeyen, P. Herdewijn, P. Marliere, *, Angew.Chemie Intern.Ed. 2013, 52, 8139-8143.*__2013__, *52*, 8139-8143.
[6] J. W. Payne, N. J. Marshall, in *Microbial Transport Systems* (Ed.: G. Winkelmann), Wiley-VCH, **2002**, pp. 139-164.
[7] T. E. Fickel, C. Gilvarg, *Nature New Biol.* **1973**, *241*, 161-163.
[8] S. M. Hammond, A. Claesson, A. M. Jansson, L. G. Larsson, B. G. Pring, C. M. Town, B. Ekstrom, *Nature* **1987**, *327*, 730-732.
[9] F. J. M. Detmers, F. C. Lanfermeijer, B. Poolman, *Res. Microbiol.* **2001**, *152*, 245-258.
[10] C. F. Higgins, *Res. Microbiol.* **2001**, *152*, 205-210.
[11] J. R. Tame, G. N. Murshudov, E. J. Dodson, T. K. Neil, G. G. Dodson, C. F. Higgins, A. J. Wilkinson, *Science* **1994**, *264*, 1578-1581.
[12] S. H. Sleigh, J. R. Tame, E. J. Dodson, A. J. Wilkinson, *Biochemistry* **1997**, *36*, 9747-9758.
[13] C. A. Guyer, D. G. Morgan, J. V. Staros, *J. Bacteriol.* **1986**, *168*, 775-779.
[14] M. K. Doeven, R. Abele, R. Tampe, B. Poolman, *J. Biol. Chem.* **2004**, *279*, 32301-32307.
[15] M. M. Klepsch, M. Kovermann, C. Low, J. Balbach, H. P. Permentier, F. Fusetti, J. W. de Gier, D. J. Slotboom, R. P. Berntsson, *J. Mol. Biol.* **2011**, *414*, 75-85.
[16] A. Marchand, D. Marchand, R. Busson, P. Marliere, P. Herdewijn, *Chem. Biodivers.* **2007**, *4*, 1450-1465.
[17] Y. J. Hu, Y. Wei, Y. Zhou, P. T. Rajagopalan, D. Pei, *Biochemistry* **1999**, *38*, 643-650.
[18] D. Chandu, D. Nandi, *Microbiology* **2003**, *149*, 3437-3447.
[19] K. Lu, Q. P. Duan, L. Ma, D. X. Zhao, *Bioconjug. Chem.* **2010**, *21*, 187-202.
[20] S.-Y. Han, Y.-A. Kim, *Tetrahedron* **2004**, *60*, 2447-2467.
[21] Y. Masui, N. Chino, S. Sakakibara, *Bull. Chem. Soc. Jpn.* **1980**, *53*, 464-468.
[22] J.-P. Mazaleyrat, J. Xie, M. Wakselman, *Tetrahedron Lett.* **1992**, *33*, 4301-4302.
[23] M. E. Jung, A. Toyota, *J. Org. Chem.* **2001**, *66*, 2624-2635.

[24] H. W. Yang, J. W. Yi, E.-K. Bang, E. M. Jeon, B. H. Kim, *Org. Biomol. Chem.* **2011**, *9*, 291-296.
[25] T. Kline, S. R. Jackson, W. Deng, C. L. M. J. Verlinde, S. I. Miller, *Nucleos. Nucleot. Nucl.* **2008**, *27*, 1282-1300.
[26] H. A. Vaccaro, Z. Zhao, J. W. Clader, L. Song, G. Terracina, L. Zhang, D. A. Pissarnitski, *J. Comb. Chem.* **2007**, *10*, 56-62.
[27] A. Semenyuk, A. Földesi, T. Johansson, C. Estmer-Nilsson, P. Blomgren, M. Brännvall, L. A. Kirsebom, M. Kwiatkowski, *J. Am. Chem. Soc.* **2006**, *128*, 12356-12357.
[28] I. Okamoto, K. Seio, M. Sekine, *Tetrahedron Lett.* **2006**, *47*, 583-585.
[29] M. Saady, L. Lebeau, C. Mioskowski, *Tetrahedron Lett.* **1995**, *36*, 2239-2242.
[30] K. C. Fylaktakidou, C. D. Duarte, A. E. Koumbis, C. Nicolau, J. M. Lehn, *ChemMedChem* **2011**, *6*, 153-168.
[31] P. Narayanasamy, D. C. Crick, *Heterocycles* **2008**, *76*, 243-247.
[32] J.-i. Matsuzaki, H. Hotoda, M. Sekine, T. Hata, S. Higuchi, Y. Nishimura, M. Tsuboi, *Tetrahedron* **1986**, *42*, 501-513.
[33] G. Dufner, R. Schwörer, B. Müller, R. R. Schmidt, *Eur. J. Org. Chem.* **2000**, *2000*, 1467-1482.
[34] M. Leuck, K. E. Vagle, J. Shawn Roach, A. Wolter, *Tetrahedron Lett.* **2004**, *45*, 321-324.
[35] K. Burgess, R. A. Gibbs, M. L. Metzker, R. Raghavachari, *J. Chem. Soc., Chem. Commun.* **1994**, *0*, 915-916.
[36] D. Hutter, M.-J. Kim, N. Karalkar, N. A. Leal, F. Chen, E. Guggenheim, V. Visalakshi, J. Olejnik, S. Gordon, S. A. Benner, *Nucleos. Nucleot. Nucl.* **2010**, *29*, 879-895.
[37] J. Guo, N. Xu, Z. Li, S. Zhang, J. Wu, D. H. Kim, M. Sano Marma, Q. Meng, H. Cao, X. Li, S. Shi, L. Yu, S. Kalachikov, J. J. Russo, N. J. Turro, J. Ju, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 9145-9150.
[38] J. P. Horwitz, J. Chua, J. A. Urbanski, M. Noel, *J. Org. Chem.* **1963**, *28*, 942-944.
[39] J. Cieslak, A. Grajkowski, C. Ausin, A. Gapeev, S. L. Beaucage, *Nucleic Acids Res.* **2012**, *40*, 2312-2329.
[40] T. W. Johnson, S. P. Tanis, S. L. Butler, D. Dalvie, D. M. DeLisle, K. R. Dress, E. J. Flahive, Q. Hu, J. E. Kuehler, A. Kuki, W. Liu, G. A. McClellan, Q. Peng, M. B. Plewe, P. F. Richardson, G. L. Smith, J. Solowiej, K. T. Tran, H. Wang, X. Yu, J. Zhang, H. Zhu, *J. Med. Chem.* **2011**, *54*, 3393-3417.
[41] E. Rozners, D. Katkevica, R. Stromberg, *Chembiochem* **2007**, *8*, 537-545.
[42] J. E. Gallant, S. Deresinski, *Clin. Infect. Dis.* **2003**, *37*, 944-950.
[43] C. Lu, Y. Jia, J. Yang, Y. Song, W. Liu, Y. Ding, X. Sun, A. Wen, *Clin. Drug Invest.* **2012**, *32*, 333-338.
[44] N. Parey, C. Baraguey, J. J. Vasseur, F. Debart, *Org. Lett.* **2006**, *8*, 3869-3872.
[45] R. Johnsson, J. G. Lackey, J. J. Bogojeski, M. J. Damha, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3721-3725.
[46] A. R. Martin, T. Lavergne, J.-J. Vasseur, F. ß. Debart, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4046-4049.
[47] P. J. L. M. Quaedflieg, C. M. Timmers, V. E. Kal, G. A. van der Marel, E. Kuyl-Yeheskiely, J. H. van Boom, *Tetrahedron Lett.* **1992**, *33*, 3081-3084.
[48] J. G. Lackey, D. Mitra, M. M. Somoza, F. Cerrina, M. J. Damha, *J. Am. Chem. Soc.* **2009**, *131*, 8496-8502.
[49] E. Kiuru, M. Ora, L. Beigelman, L. Blatt, H. Lönnberg, *Chem. Biodivers.* **2011**, *8*, 266-286.
[50] W. Shen, J.-S. Kim, P. E. Kish, J. Zhang, S. Mitchell, B. G. Gentry, J. M. Breitenbach, J. C. Drach, J. Hilfinger, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 792-796.
[51] M. G. Friedel, J. C. Pieck, J. Klages, C. Dauth, H. Kessler, T. Carell, *Chem. Eur. J.* **2006**, *12*, 6081-6094.
[52] K. S. Ghosh, J. Debnath, P. Dutta, B. K. Sahoo, S. Dasgupta, *Bioorg. Med. Chem.* **2008**, *16*, 2819-2828.
[53] H. Cui, L. M. Ruiz-Perez, D. Gonzalez-Pacanowska, I. H. Gilbert, *Bioorg. Med. Chem.* **2010**, *18*, 7302-7309.

Chapter 3

Synthesis of novel 5'-linked peptide nucleotide conjugates (PNCs) for nucleotides delivery in bacterial cells

ABSTRACT

The rational design and synthesis of novel peptide nucleotide conjugates (PNCs), assembled from α-substituted glycine dipeptides and TMP, are described. The two composing structural units are connected by various phosphoester linkers between the α-carbon of a C-terminal glycine residue and the 5'-hydroxyl group of thymidine. The resulting conjugates are of interest in the development of a nucleotide delivery system in prokaryotes and as potential artificial substrates for DNA polymerases. The performance of all dipeptide-TMP conjugates as pyrophosphate mimics and the influence of the different linkers in the DNA polymerization reaction were evaluated by *in vitro* chain elongation assay, using both thermophilic and mesophylic microbial DNA polymerases.

3.1. INTRODUCTION

In **Chapter 2** we have already discussed the relevance of peptide transport systems[1] for potential nucleotide delivery in bacterial cells, detailing the design of peptide-nucleotide conjugates (PNCs), which were prepared by covalently binding, through various linkers, the side chain functionalities of oligopeptides (di- to penta-) to the 3'-position of thymidine monophosphate.

Throughout our search for a suitable peptidic-based system for cargo delivery in bacterial cells, we came across a report by Gilvarg *et al*., which described glycine-containing peptides comprising a nucleophilic moiety or a toxophoric agent (mostly uncharged small molecules) attached to the α-carbon of the C-terminal glycine residue.[2] Such peptide-conjugates were found not only to be accepted by peptide transporters, but also to undergo hydrolysis in the presence of intracellular peptidases, thus liberating the attached cargo inside the bacterial cells (Fig 3-1).[3] Inspired by this model, we envisioned that this type of peptide delivery system could also be functional for nucleotide delivery. In order to test this hypothesis, we chose to synthesize 5'-thymidylate (TMP) nucleotide conjugates with various α-substituted glycine dipeptides as initial synthetic targets, for evaluation within our well-established nutritional selection system, using auxotrophic *E. coli* mutants lacking thymidylate synthase (Fig 3-2). These peptide 5'-nucleotide conjugates (PNCs) were thought to be embedded with the potential to liberate TMP intracellularly based on their variable metabolic (in)stability, depending on the different phosphate, phosphoramidate, phosphorothioate and thioethyl-phosphate ester functionalities used to connect the α-substituted glycine dipeptide (L-Ala-Gly) to the 5'-position of thymidine.

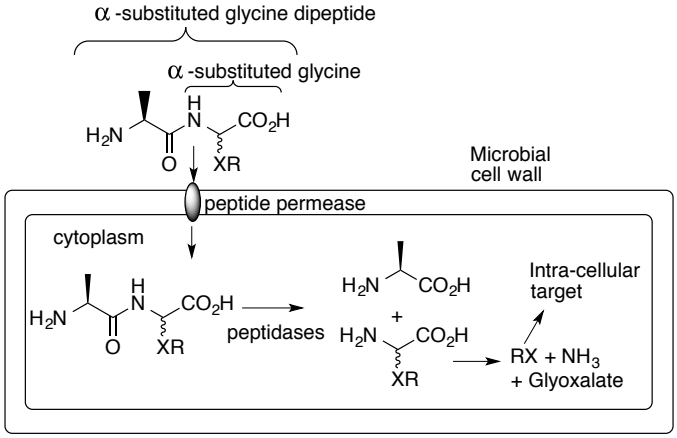


Fig 3-1: Structures of α-substituted glycine peptide-conjugates and their mode of breakdown after peptidase cleavage.

α-Substituted glycines are highly reactive compounds, but their stabilization can be accomplished by masking the lone pair of the glycyl nitrogen through N-acylation.[4] Due to their inherent chemical instability, however, the synthesis of the corresponding PNCs cannot be accomplished by standard methods. Generally, the α-substituent in α-substituted glycines acts as a good leaving group, therefore

the synthesis of α-substituted phosphoester glycines is more challenging due to the likely increase of nucleophilicity at the α-carbon.

Although a number of reports illustrate the synthesis of carba(cyclic) and simple hetero-atomic (aromatic)[2, 3, 5] α-substituted glycines, careful literature investigation revealed the lack of synthetic methods for the introduction of a phosphoester group at the α-carbon of glycines. Herein, we report the first syntheses of PNCs of α-substituted glycines with different phospho-ester groups, such as phosphate, phosphoramidate, phosphorothioate and thioethyl-phosphate.

3.2. RESULTS and DISCUSSION

Phosphorus-containing α-substituted glycines pose a greater synthetic challenge when compared to uncharged hetero-atomic α-substituted glycines. Synthetic routes have to be carefully designed and executed, resorting to suitable orthogonal protection-deprotection methods, to yield labile phosphoester moieties. Our retrosynthetic analysis for the proposed phospho-ester target compounds is shown in Fig. 3-3. It becomes apparent that those conjugates can be obtained by nucleophilic substitution either at the α-position of glycine (route 1), or at the phosphorus group itself (route 2).

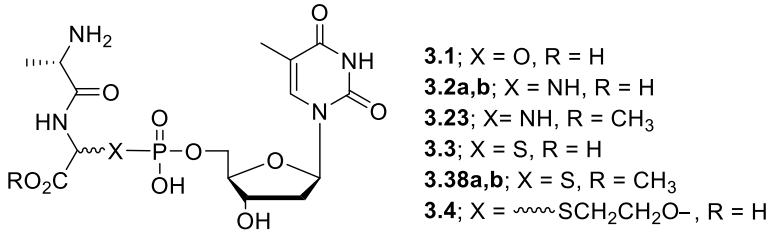


Fig 3-2: Structures of α-substituted glycine peptide-nucleotide conjugates with different phosphoester linkages.

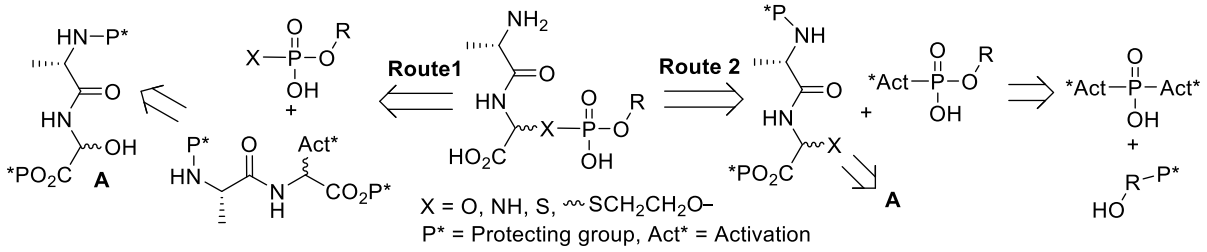
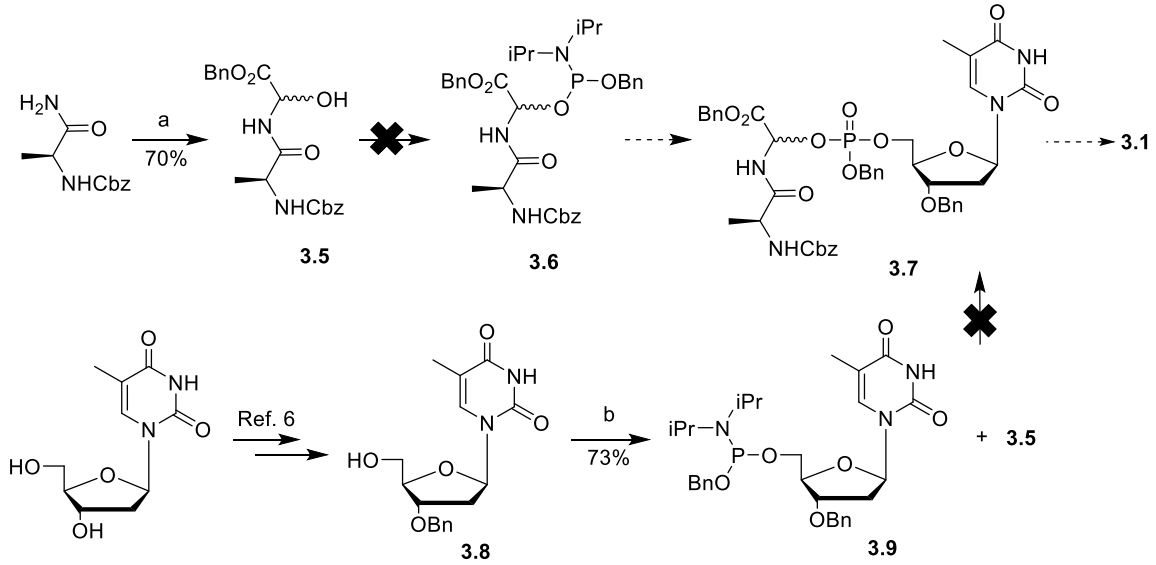


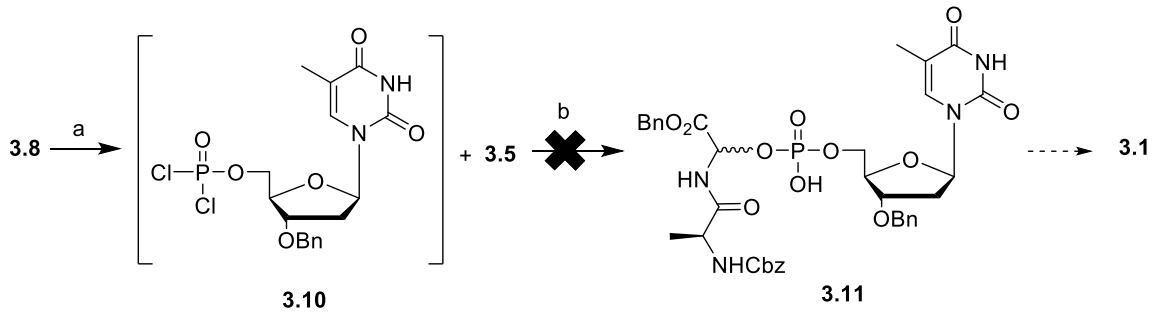
Figure 3-3. Retrosynthetic analysis of planned PNCs.

Among the various options available for the formation of a phosphodiester linkage, the approach based on the use of a dialkylphosphoramidite followed by *in situ* oxidation was selected in our initial efforts to access phosphate derivative **3.1**. α-Hydroxy glycine derivative **3.5** was synthesized as a diastereomeric mixture in good yield, after coupling Z-L-Ala-NH_2 with benzyl glyoxalate according to an improved protocol, including the addition of a catalytic amount of p-TSA to increase the reaction

rate.[2] In order to access the target molecule **3.1**, we first attempted to make bis(alkyl)-*N,N*-diisopropyl phosphoramidite **3.6**, but as the reaction met with failure, we turned to an alternative strategy by converting 3-*O*-benzyl thymidine[6] **3.8** into the pro-moiety **3.9**,[7] followed by phosphoramidite coupling with **3.5** and in situ oxidation. Although phosphoramidite **3.9** could be successfully isolated, the subsequent coupling with **3.5** in the presence of 1-*H* tetrazole and oxidation with H_2O_2 did not produce compound **3.7**, but 3'-*O*-benzyl-(benzyloxy)-TMP was obtained instead. Despite various coupling agents with different pK_a, like 4,5-dicyanoimidazole, 2-ethylthio tetrazole and 2-benzylthio tetrazole were screened in the phosphoramidite coupling with the α-OH of **3.5**, none of those conditions were found to be beneficial.



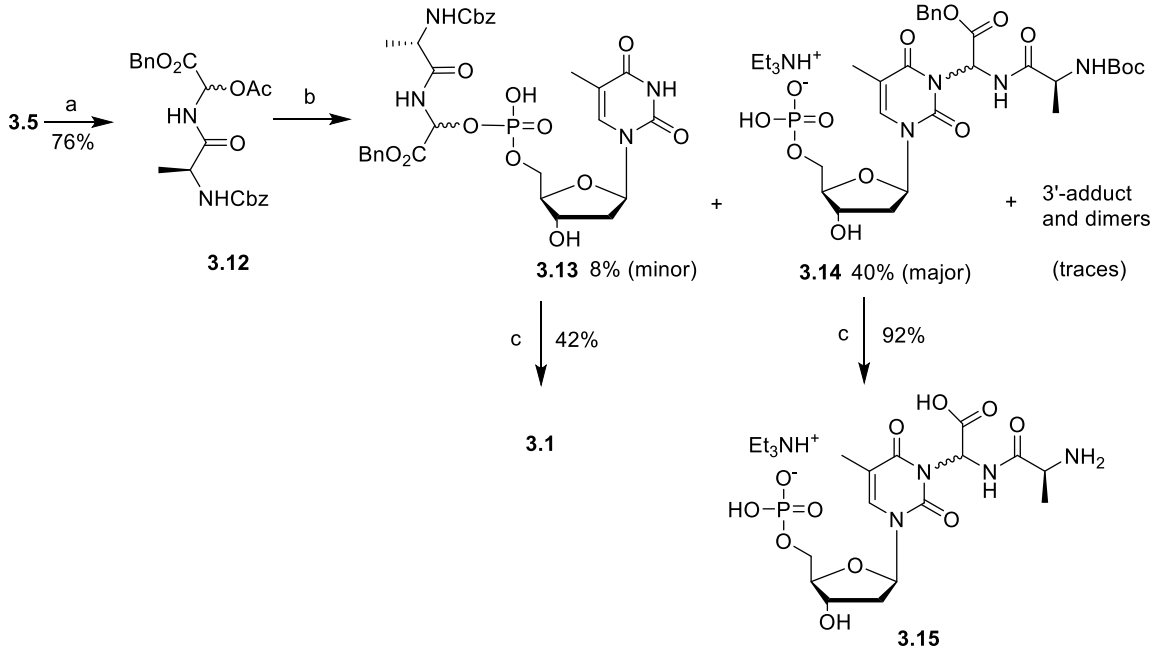
Scheme 3-1. Route towards the synthesis of 5'-*O*-(L-alanyl-D,L-2-aminoglycine)-TMP. Reagents and conditions: (a) benzyl glyoxalate, p-TSA, CH_2Cl_2, 168 h; (b) Benzyloxy-bis(*N*,*N*-diisopropylphosphoramino) phosphine, 0.45 M 1*H*-tetrazole in CH_3CN, CH_2Cl_2, 0 °C to r.t., 4 h.



Scheme 3-2. Route towards the synthesis of 5'-*O*-(L-alanyl-D,L-2-aminoglycine)-TMP. Reagents and conditions: (a) $POCl_3$, trimethyl phosphate, -20 °C to r.t., 8 h; (b) Base, CH_2Cl_2, 0 °C to r.t., 16 h.

As the phosphoramidite approach was not successful to provide phosphate derivative **3.1**, we decided to use a method based on a more reactive chlorophosphate intermediate, as described in Scheme 2. First, 3'-*O*-benzyl thymidine **3.8** was converted to its active form 5'-*O*-dichlorophospho-3'-*O*-benzyl-thymidine **3.10** using $POCl_3$ in trimethyl phosphate, which was then treated *in situ* with

3.5 in the presence of *N*-methyl imidazole. Surprisingly, the reaction did not afford the desired product, but resulted in the formation of 3-*O*-benzyl-TMP along with other uncharacterized by-products. Since bases with high pK_b are known to form acyliminio glycine intermediates,[8] a number of weaker bases such as Et_3N, DIPEA and pyridine were employed to increase the basicity of the α-OH of glycine, but yielded almost the same results. This may be due to the lower nucleophilicity of the α-OH group, a sterical hindrance at the α-position, or the instability of the mono-chlorophosphate intermediate during quenching of the reaction.

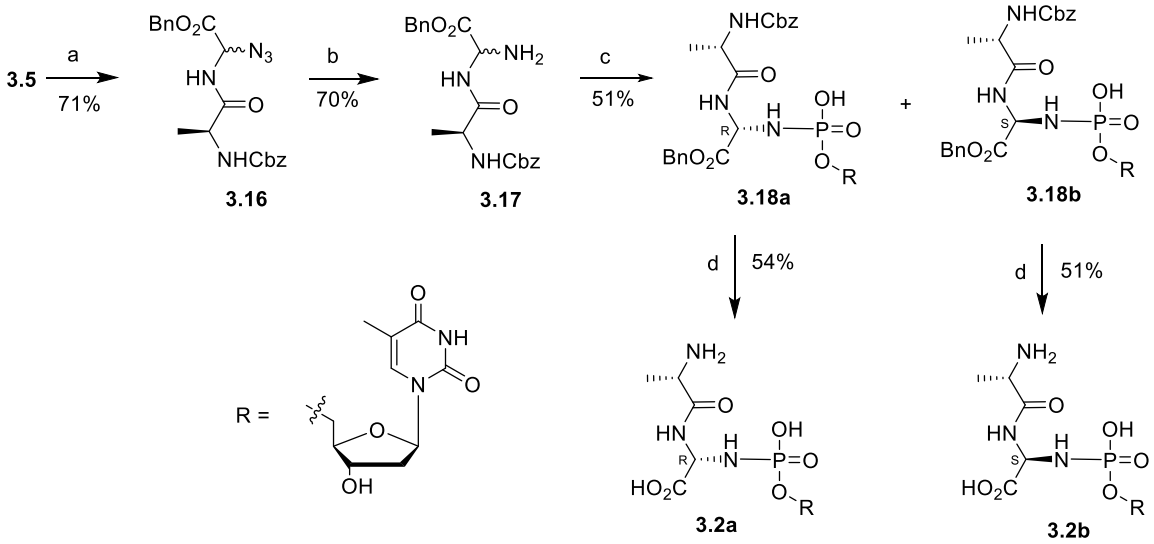


Scheme 3-3. Synthesis of 5'-*O*-(L-alanyl-D,L-2-aminoglycine methyl ester)-TMP. Reagents and conditions: (a) Ac_2O, pyridine, 0 °C, 8 h; (b) TMP triethylammonium salt, Et_3N, DMF, 40 °C, 24 h; (c) 10% Pd/C (Degussa-type), EtOH:H_2O 5:1, r.t., 4 h.

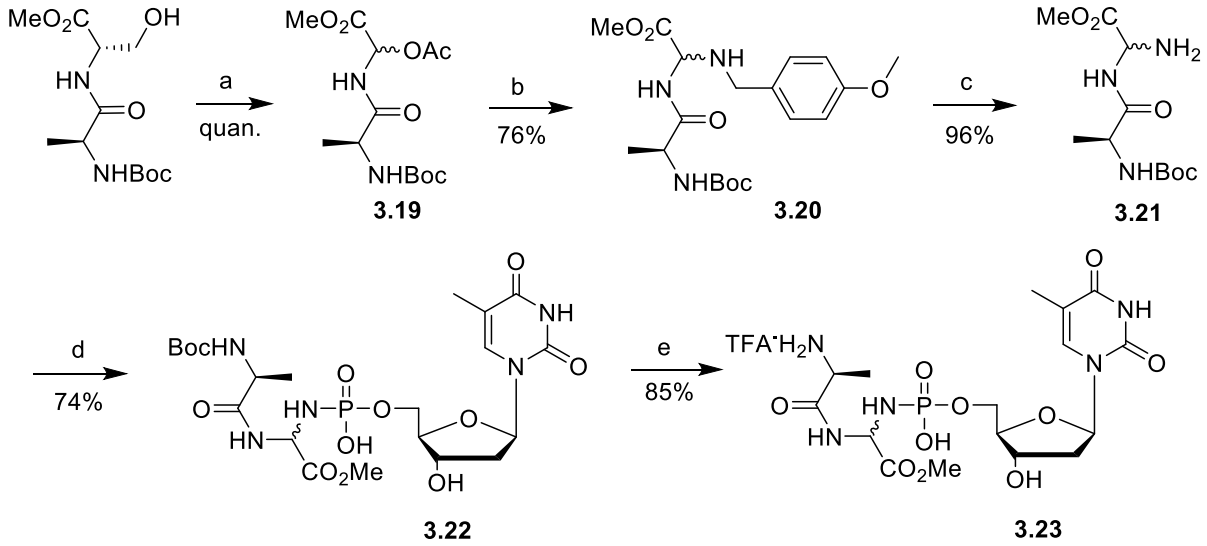
Those unsatisfactory results led us to switch to Route 1 in the retrosynthetic analysis for the synthesis of **3.1**. Thus, the α-hydroxy glycine derivative **3.5** was activated at the α-position for nucleophilic substitution by conversion to its α-acetoxy derivative **3.12**, which was then reacted with the triethylammonium salt of TMP (Scheme 3-3). Although at first a single product was thought to have formed on the basis of mass spectroscopic analysis of the crude residue, results from RP-HPLC purification revealed a mixture of products with the same mass. After in-deepth 2D NMR analysis of the separated products, it could be concluded that the desired phosphodiester derivative **3.13** was obtained as a minor product (8%) along with thymidylate N^3-peptide adduct **3.14** as major product (40%). A 3'-TMP-peptide adduct and different dimers were also produced in traces in the reaction. This result can be rationalized on the basis of the hard-soft acid base (HSAB) principle, according to which a α-carbon, being a soft center, prefers a soft donor like the electrons from the N^3 atom rather than the oxygen atom of the phosphate group. Despite attempts to optimize the reaction yield, longer times and higher temperatures only led to the formation of more undesired side-products. As the 3'-

TMP-peptide adduct and compound **3.13** have very similar R_f, it was difficult to separate the two regio-isomers even by preparative-HPLC, thus the compounds were eventually separated in the next step. Finally, removal of all protecting groups of **3.13** by reductive hydrogenation using 10% Pd/C, Degussa-type, as previously optimized in our group, and preparative HPLC purification yielded the desired PNC **3.1** in moderate yield (42%). During the conversion of **3.13** to **3.1**, the formation of TMP as an impurity was observed, whilst the same reaction conditions applied to **3.14** gave **3.15** in excellent yield (92%).

The synthesis of phosphoramidate analogue **3.2** could theoretically be accessed either via Route 1 or 2 as shown in the retrosynthetic analysis (Fig 3-3). As earlier witnessed in Route 1, a potential competition for nucleophilicity might arise between the N^3-atom of thymine and the *N*-atom of thymidine-5'-O-phosphoramidate, therefore the more selective approach Route 2 was employed. The synthesis of α-amino glycine derivative **3.17** from its corresponding hydroxyl derivative **3.5** was performed via azide intermediated conversion (Scheme 3-4). Azide **3.16** was prepared from **3.5** in good yield using the DPPA/DBU method[9] and it was then converted to its amino form **3.17** by classical azide reduction in the presence of PPh_3. The phosphoramidate coupling was carried out under standard DCC coupling conditions[10] to obtain a diastereomeric mixture of compounds **3.18a** and **3.18b** in good yield. Our initial efforts to separate the isomers **3.18a-b** after the final debenzylation step turned out to be problematic, however the protected diastereomers could be successfully separated by preparative RP-HPLC yielding two pure **3.18a** and **3.18b** in an approximate 1:1 ratio. Finally, the protecting groups were removed by catalytic hydrogenation as described earlier to afford the corresponding isomers of desired phosphoramidates **3.2a** and **3.2b** in moderate yields.



Scheme 3-4. Synthesis of 5'-*O*-(L-alanyl-D/L-2-aminoglycine methyl ester)-TMP **3.2a** and **3.2b**. Reagents and conditions: (a) DPPA, DBU, dry toluene, 0 °C to r.t., 16 h; (b) PPh_3, THF, reflux, 2 h; (c) TMP triethylammonium salt, Et_3N, DCC, ^{t}BuOH:H_2O 4:1, 85 °C, 2.5 h; (d) 10% Pd/C (Degussa-type), EtOH:H_2O 5:1, r.t., 4 h.

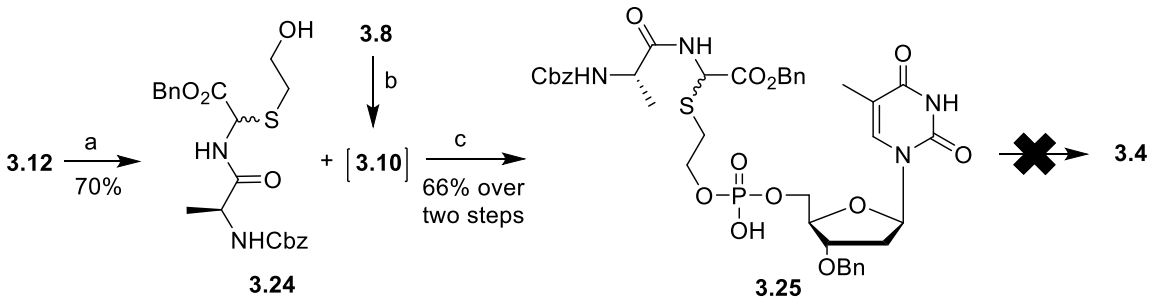


Scheme 3-5. Synthesis of thymidine-5'-(L-alanyl-D,L-2-aminoglycine methyl ester) phosphoramidate. Reagents and conditions: (a) $Pb(OAc)_4$, dry EtOAc, mol. sieves 4 Å, reflux, 2 h; (b) 4-methoxybenzylamine, DIPEA, dry DMF, 40 °C, 24 h; (c) 10% Pd/C, EtOH, r.t., 5 h (d) TMP triethylammonium salt, DCC, Et_3N, ^{t}BuOH:H_2O 4:1, 85 °C, 2.5 h; (e) TFA, thioanisole, CH_2Cl_2, 0 °C to r.t., 4 h.

Since it can be anticipated that the free α-carboxylic group of glycine might play a role in the formation of TMP as a side product during the hydrogenolysis of **3.13**, **3.18a** and **3.18b**, we were interested to synthesize phosphoramidate analogue **3.23**, where the carboxylic group of glycine is protected as a methyl ester. A suitable choice of orthogonal protecting groups was crucial for the synthesis of phosphoramidate **3.23** (Scheme 3-5), whilst the main synthetic strategy resembled that employed for **3.2a-b**. Unlike benzyl-protected α-acetoxy glycine derivative **3.12** which was synthesized in a convergent manner, for the synthesis of α-actoxy derivative **3.19** a more practical degradative approach was used, starting from *N*-Boc-L-Ala-L-Ser-OMe in the presence of $Pb(OAc)_4$,[8] as the protecting groups were well tolerated during the reaction. The acetoxy derivative **3.19** was then converted to the protected amino derivative **3.20** by nucleophilic substitution with 4-methoxybenzyl amine. Subsequent deprotection by Pd/C catalyzed hydrogenation yielded the desired amine **3.21** in excellent yield (96%). Phosphoramidate **3.22** was obtained in good yield using the standard protocol and removal of the Boc group with TFA in the presence of the radical scavenger thioanisole produced the desired methyl ester derivative **3.23**. As for compounds **3.18a-b**, a late stage effort was made to separate the diastereomers of **3.22** and **3.23** by preparative RP-HPLC, but was not effective in this case.

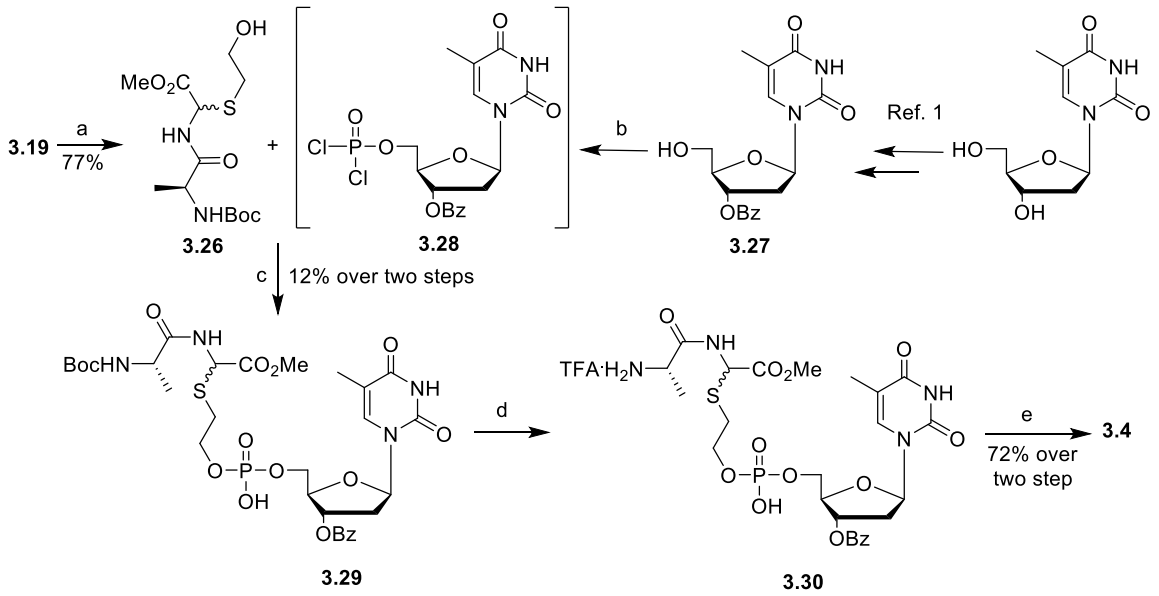
Although both retro-synthetic routes are feasible for the synthesis of thioethyl phosphate derivative **3.4**, the instability of the thioethyl-TMP synthon can be considered potentially problematic. As described in the literature, this unit can be readily synthesized[11] but as soon as it is formed from its precursor, it immediately decomposes with formation of TMP and thiirane via nucleophilic attack of the naked thiol group.[12] According to Route 2, thioethanol derivative **3.24** derived from the

corresponding acetoxy derivative **3.12** and 2-mercaptoethanol by selective S-alkylation in good yield (70%) (Scheme 3-6). The phosphoramidite approach could in theory be utilized for the synthesis of **3.25**, nonetheless selective oxidation of P^{III} to P^{V} in the presence of a sulfur atom pose a real challenge.[13] Therefore by alternatively adopting the $POCl_3$ method, intermediate **3.10** was obtained from the corresponding 3'-*O*-benzyl thymidine, and it was then treaded in situ with **3.24** in the presence of *N*-methyl imidazole. Quenching followed by preparative RP-HPLC purification furnished **3.25** as a diastereomeric mixture, in 66% yield over two steps. For the conversion of **3.25** to **3.4**, removal of the Cbz and benzyl groups by hydrogenation was performed using a range of palladium derivatives, including 10% Pd/C, 10% Pd/C (Degussa), 20% $Pd(OH)_2/C$, Pd-black and $Pd(OAc)_2$. The use of an excess of catalyst was necessary in view of catalyst poisoning occurring in the presence of sulfur atoms. However, all catalysts failed to provide **3.4**, either producing undesired over-reduced side products under harsher conditions or starting material under milder conditions.



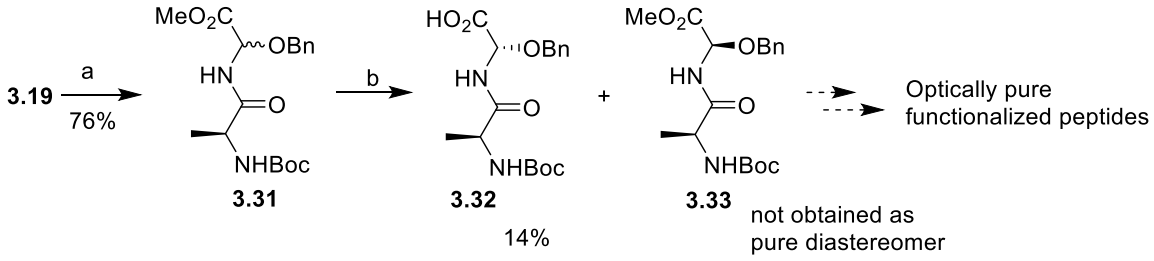
Scheme 3-6. Route towards the synthesis of 5'-*O*-(L-alanyl-D,L-2-mercaptoethanolglycine) TMP. Reagents and conditions: (a) 2-mercaptoethanol, Et_3N, dry DMF, 0 °C to r.t.,16 h; (b) $POCl_3$, trimethyl phosphate, -20 °C to r.t., 8 h (c) *N*-methyl imidazole, CH_2Cl_2, 0 °C to r.t.,16 h.

It was therefore decided to turn to an alternative protecting group approach for the synthesis of **3.4** (Scheme 7). Thioethanol derivative **3.26** was isolated in good yield (77%) from its acetoxy derivative **3.12**. Analogously to the synthesis of **3.25**, the corresponding phospho-diester derivative **3.29** was obtained in low yield (12%) over two steps. It was found that the low conversion rates of **3.27** into intermediate **3.28** is the main cause for the formation of complex by-products in the next base-promoted step, due to the presence of unreacted starting material. Other methods were explored in order to improve the yield of **3.28**, including the use of $POCl_3$ with different bases and solvents,[14] but unfortunately none led to improved conversions, and the formation of dimers was detected even under controlled conditions. Finally, the removal of the Boc group was achieved via the TFA/Thioanisole method, followed by the simultaneous hydrolysis of glycine-ester and the 3'-benzoyl group with LiOH, affording **3.4** in good yield (72%) over two steps.



Scheme 3-7. Synthesis of 5'-*O*-(L-alanyl-D,L-2-mercaptoethanolglycine) TMP. Reagents and conditions: (a) 2-mercaptoethanol, Et_3N, dry DMF, 0 °C to r.t.,16 h; (b) $POCl_3$, trimethyl phosphate, -20 °C to r.t., 8 h (c) *N*-methyl imidazole, CH_2Cl_2, 0 °C to r.t.,16 h; (d) TFA, CH_2Cl_2, thioanisole, 0 °C to r.t., 4 h; (e) LiOH, MeOH:H_2O 1:1, 0 °C to r.t., 3 h.

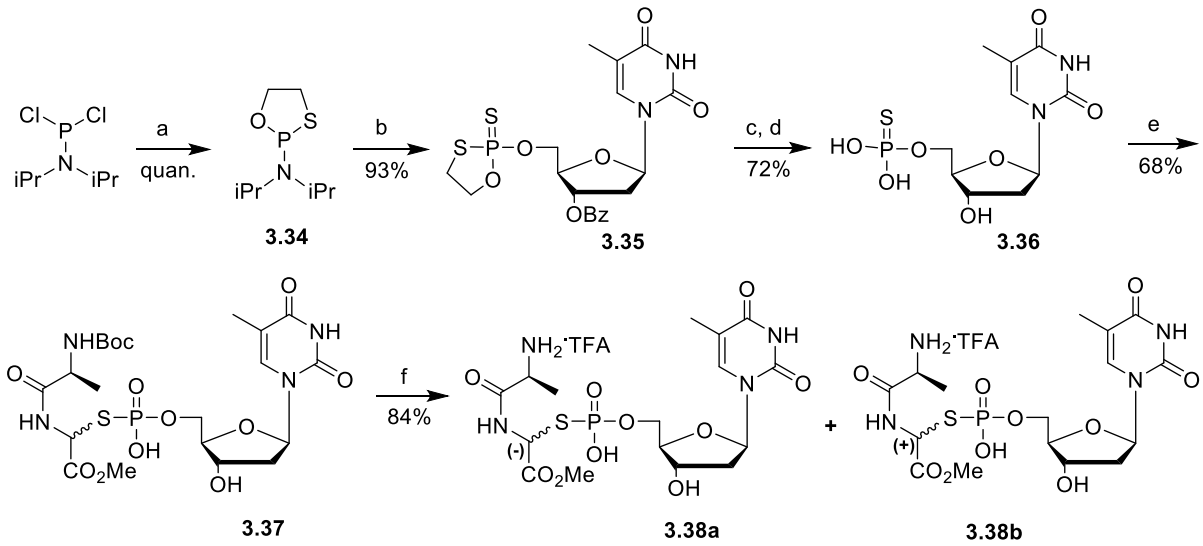
As separation of diastereomers by preparative RP-HPLC represents a limitation with regard to the quantities of products that can be purified, we tried to perform this step at the latest stage of the synthetic pathway for all the conjugates. Moreover effective separation could be achieved only in some cases. Therefore it was thought that the enzymatic separation of starting diastereomeric peptides could offer a solution to obtain large quantities of pure final compounds. Accordingly, the diastereomeric mixture of *N*-Boc-L-Val-D,L-benzyloxy-Gly-OMe was successfully resolved by enzyme catalyzed ester hydrolysis using subtilin Carlsberg (Scheme 3-8).[8] Conversely, similar conditions applied to *N*-Boc-L-Ala-D,L-benzyloxy-Gly-OMe resulted in a lousy separation, may be due to poor substrate specificity.



Scheme 3-8. Route towards the enzymatic separation of diastereomers. Reagents and conditions: (a) DABCO, dry THF, BnOH, -78 °C to r.t., 24 h; (b) Subtilisin Carlsberg, 1M NaOH, DMF/H_2O 1:1, pH = 7.2, 55 °C.

For our last targets, phosphorothioate derivatives **3.38a** and **3.38b**, the initially followed retro-synthetic Route 2 was replaced by the alternative approach described in Scheme 3-9, as it was not

possible to isolate the *N*-Boc-L-alanyl-D,L-2-thioglycine methyl ester synthon in pure form, likely due to the instability of the free thiol group. The synthesis of thymidine-5'-*O*-phosphorothioate **3.36** is reported in the literature via a solid-phase methodology.[15] Despite being effective for small amounts, on scale-up this protocol was not reproducible. Thus, monothiophosphorylation with $PSCl_3$[16] was next explored to access **3.36** in one step from thymidine, but the reaction rates were found to be sluggish and the process non-selective in the presence of an excess of reagent. However, the protection of 3'-OH of thymidine and the use of an improved oxathiaphospholane approach adapted from the literature were, could overcome the low conversions.[17] As depicted in Scheme 3-9, compound **3.34** was obtained from *N,N*-diisopropyl phosphoramidous dichloride and 2-mercapto ethanol in the presence of DIPEA, and, without purification, was reacted with 3'-*O*-benzoyl thymidine and S-ethylthio tetrazole. In situ sulfurization with elemental S_8 afforded **3.35** in excellent yield (93%) over three steps. Opening of the oxathiaphospholane ring of **3.35** with 3-hydroxypropionitrile and simultaneous removal of 2-cyanoethyl and 3'-benzoyl groups using 25% aqueous ammonia in a sealed tube produced **3.36** in good yield. Compound **3.37** was obtained in just 4 h as a diastereomeric mixture following selective S-alkylation over N^3-alkylation, by reacting compound **3.36** with compound **3.19**. Ultimately Boc cleavage and preparative RP-HPLC purification allowed the isolation of diastereomers **3.38a** and **3.38b**. When the two diastereomers were separately subjected to ester hydrolysis with LiOH , TMP was mainly isolated.



Scheme 3-9. Synthesis of thymidine-5'-*O*-(L-alanyl-D,L-2-aminoglycine methyl ester)-phosphorothioate. Reagents and conditions: (a) 2-mercapto ethanol, DIPEA, diethyl ether, -20 °C to r.t., 2.5 h, (b) (i) 3'-*O*-benzoyl thymidine, *S*-ethylthio tetrazole, CH_2Cl_2, 0 °C to r.t., 3 h; (ii) S_8, r.t., overnight; (c) 3-hydroxy propionitrile, DBU, r.t., 8 h; (d) 25% NH_3, 55 °C, sealed tube, 16 h; (e) Compound **3.19**, Et_3N, DMF, 37 °C, 4 h; (f) TFA, thioanisole, CH_2Cl_2, 0 °C to r.t., 3 h.

The *in cellulo* mode of breakdown of α-substituted glycine peptide-conjugates after peptidase cleavage was postulated by Gilverg *et. al.*[2] In this context, the formation of TMP as side-product

during the synthesis of compounds **3.1-3.3** could be explained with the anchimeric assistance of the α-carboxylic group of glycine to the phosphorus atom leading to the formation of a 5-membered ring intermediate. This phenomenon has previously been noticed at different rates in the presence of phosphate, phosphoramidate and phosphorothiate moieties.[18] By masking the free α-carboxyl group of glycine as methyl ester it was possible to resolve the meta-stability of the corresponding compounds.

3.3. CHAIN ELONGATION EXPERIMENTS (performed by Lia Margamuljana)

As dipeptides are comparable in terms of size to the pyrophosphate group,[19] it was envisioned that L-Ala-Gly could also act as a alternative leaving group in the enzyme catalyzed DNA polymerization reaction. In this study, the activities of different thermophilic and mesophilic microbial DNA polymerases were compared, comprising Therminator, Vent (exo-) and the Klenow fragment (exo-) of *E. coli* DNA Polymerase I. In order to investigate the strand elongation capacity of all the final analogues exhibiting α-substituted glycines connected through various phospho-linkers (**3.1**, **3.2a**, **3.2b**, **3.23**, **3.4** and **3.38a-b**), a polyacrylamide gel-based template-dependent incorporation assay of multiple nucleotides was carried out, using a tagged 5'-radiolabelled γ-^{33}P or a fluorescent DNA primer (**P1**) annealed in turn to **T1-3** templates (**Table 3-1**).

Table 3-1. Overview of primer and templates sequences used in the incorporation experiments.

Primer P1: 5' /5Cy5/CAGGAAACAGCTATGAC 3'
Template T1: 3' GTCCTTTGTCGATACTGAAAAA 5'
Template T3: 3' GTCCTTTGTCGATACTGATTTT 5'

Amongst all the compounds screened as potential triphosphate mimics, phosphoramidate analogues **3.2a** and **3.2b** exhibited interesting incorporation efficiencies with all tested polymerases (Fig. 3-4A). Compound **3.1** showed poor incorporation efficiency and exclusively with Therminator polymerase. On the basis of these observations, it can be deduced that the free carboxylic group along with an appropriate chain length are both crucial factors for incorporation. Vent (exo-) showed poorer efficiency than Therminator and the Klenow fragment, producing only 4% (P + 4) strand formation for **3.2a**, and 11% (P + 3) for **3.2b** at 1mM concentration after 60 min. Therminator gave 76% full length product (P + 5) and up to 4% (P + 4) strand formation for **3.2a** and **3.2b** respectively. The Klenow fragment showed the best incorporation efficiency, leading to the formation of 27% full length product (P + 5) and up to 11% of (P + 4) strand at 1mM concentration after only 15 min. Notably, **3.2a** exhibited better substrate properties than its diastereomer **3.2b**, probably because of

different binding efficacy in the enzyme pocket which might be clarified with the aid of a modeling study.

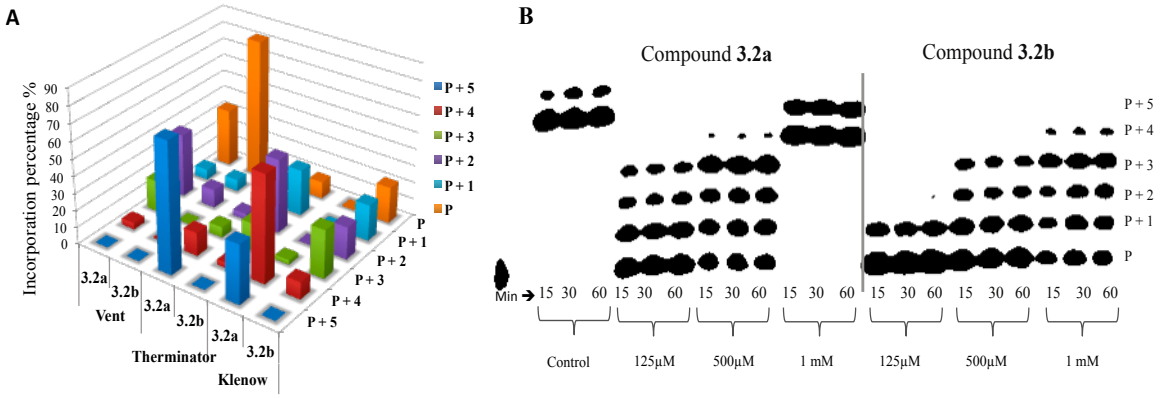


Figure 3-4. (A) Chain elongation efficiencies of compounds **3.2a** and **3.2b** by Vent (exo-) (0.01 U.µl-1), Therminator (0.01 U.µl-1) and Klenow fragment (0.05 U.µl-1) polymerases, at 1mM conc. after 60 min. (B) Chain elongation profile of control **TTP** and compounds **3.2a** and **3.2b** as substrates into the P1:T1 complex by Klenow polymerase (0.05 U.µl-1).

Table 3-2 Steady–state kinetics of single nucleotide incorporation into P1T3 by Klenow fragment (0.005 U.µl-1).

Substrate	V_{max} [nM.min^{-1}]	K_m [µM]	V_{max} / K_m [x 10^{-3} min^{-1}]
TTP	30.59 ± 0.74	0.088 ± 0.006	347.61
Compound **2a**	30.48 ± 1.04	87.21 ± 9.21	0.3495

As the best incorporation results were obtained with phosphoramidate **3.2a** and Klenow fragment (exo-) polymerase, the kinetic parameters of **3.2a** were compared with its natural substrate counterpart TTP, based on the Single Complete Hot model. Steady-state kinetic analysis (Table 3-2) indicates a comparable V_{max} value, but a K_m which is 991-fold higher, with a final V_{max}/K_m ratio for compound **3.2a** 994-fold lower than the natural substrate. This result is falls within the range of kinetic values previously reported for other leaving groups in our lab.[10]

3.4. MODELING STUDY OF 3.2a AND 3.2b BOUND TO *E. coli* KLENOW FRAGMENT (DNA polymerase) (performed by Prof. Mathy Froeyen)

Based on a homology model of *E. coli* Klenow fragment of DNA polymerase I (details can be found in the experimental section and supplementary information), both diastereomers **3.2a** and **3.2b**

were able to bind in the enzyme-binding site for the accommodation of triphosphates. The molecular mechanics energy of the complex enzyme-DNA duplex-**3.2a** is slightly lower than that of the similar complex with **3.2b**. The better energetic profile for the **3.2a** molecule complex can be attributed to the extra ionic interaction of the carbonyl group with one Mg^{2+} ion in the **3.2a**-complex. In the **3.2b**-complex, the Mg^{2+} ion interacts with an uncharged carbonyl group of the peptide bond, which is supposed to be weaker. Additionally, **3.2a** forms also a hydrogen bond between this C=O and the residue I386 backbone nitrogen and possibly a stabilizing cationic-pi-interaction of the charged NH_3^+ to the F439 electron rich aromatic ring. This modeling study clearly indicates that a stronger binding affinity to the enzymatic site might be the key factor for the better incorporation efficiency observed for **3.2a** over **3.2b**. The determination of the absolute configuration (R/S) of each diastereomer is based on incorporation efficiency and modeling study.

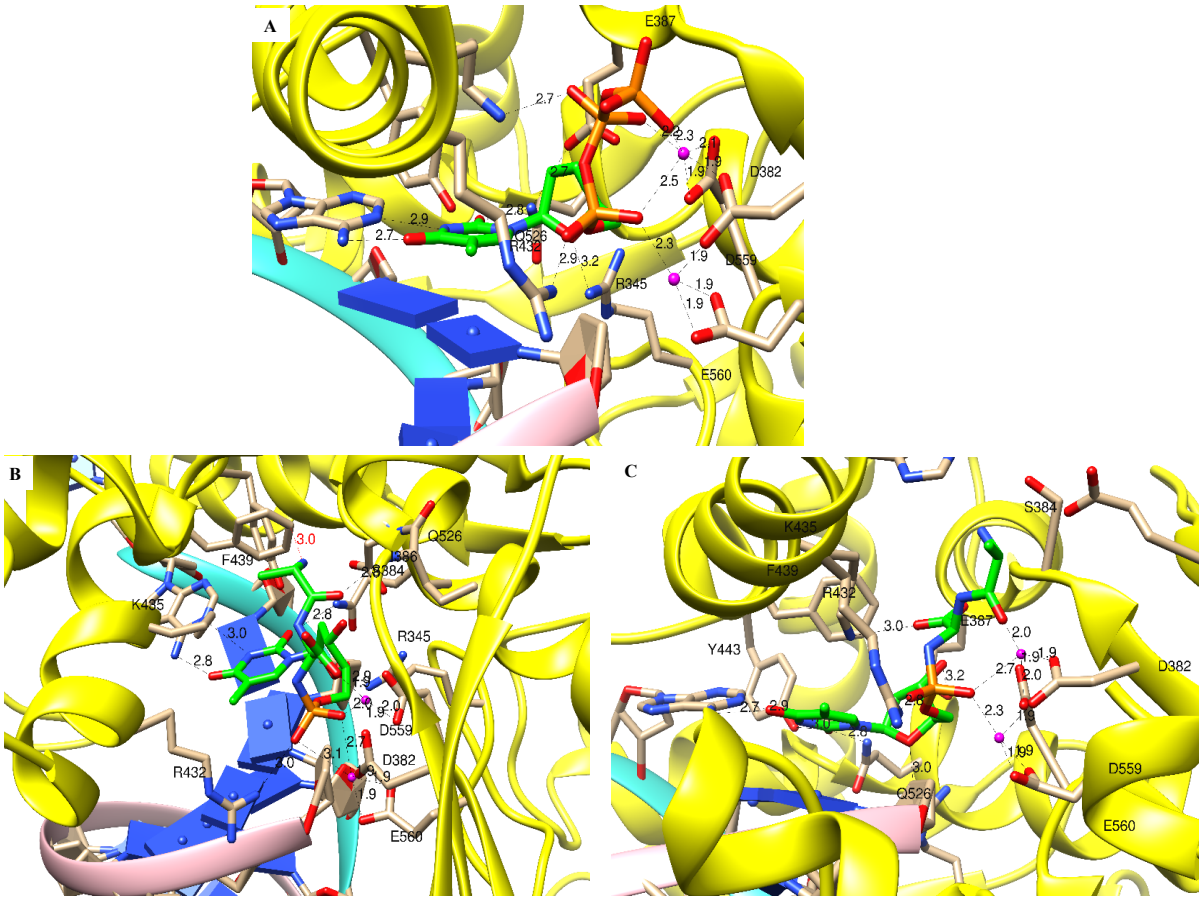


Figure 3-5. Compounds **TTP** (Fig 3-5A), **3.2a** (Fig 3-5B) and **3.2b** (Fig 3-5C) in the *E. coli* Klenow fragment pocket. The primer strand is shown as a cameo pink ribbon, whilst the template strand is in cyan. Hydrogen bond distances between the bases are indicated in Å. The ligation of the $2Mg^{2+}$ ions to the complex is also visualized by dashes and distances in Å. The figures were generated using chimera.

3.5. CONCLUSION

A novel series of 5'-modified derivatives of TMP with the dipeptide L-Ala-Gly were successfully prepared, by nucleophilic substitution either at the α-position of glycine or at the phosphorous-group of TMP. Different phosphoester linkers such as phosphate, phosphoramidate, phosphorothioate and thioethyl-phosphate were used to connect those PNCs. In a template-based multiple incorporation assay, analogues **3.2a** and **3.2b** featuring a phosphoramidate linker and the Klenow fragment (exonuclease-free) of DNA polymerase I were found to be the best substrate and the most effective polymerase respectively, giving the best strand elongation rates. A free carboxylic group at the C-terminal of the dipeptidic chain was necessary for binding in the active site of the enzyme, and is also accountable for the destabilization of the conjugates resulting from hydrolytic cleavage through anchimeric assistance. The further exploration of such dipeptides as leaving groups for DNA polymerization can be a valuable tool for synthetic biology, drug delivery, polymerase chain reaction (PCR) and as potential prodrugs for antiviral therapy.

3.6. EXPERIMENTAL SECTION

General Information

For all reactions, analytical grade solvents were used. TMP-disodium salt was transformed into its triethylammonium salt, first by treatment with Amberlite IRA-120B ion exchange resin (H^+ form) and then with triethylamine. All moisture-sensitive reactions were carried out using oven-dried glassware (135 ºC) under a nitrogen or argon atmosphere. Reaction temperatures are reported as bath temperatures. Pre-coated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ = 254 nm). Products were purified by flash chromatography on ICN silica gel 63-200, 60 Å. All final compounds were purified by preparative RP-HPLC (Xbridge™ Prep C18 5µm OBD 19 x 150 mm column) using a gradient of H_2O and CH_3CN, or both containing 50 mmol TEAB as eluent buffer. ^1H, ^{13}C and ^{31}P NMR spectra were recorded on Bruker Avance 300 MHz, 500 MHz or 600 MHz spectrometers. Final compounds were characterized using 2D NMR (H-COSY, HSQC, HMBC, TOCSY and NOESY) techniques. For sake of clarity, NMR signals of protons and carbons for sugar and base moieties are indicated with and without a prime, respectively. Chemical shifts were referenced to residual solvent signals at δ H/C 7.26/77.00 ($CDCl_3$), 3.31/49.10 (CD_3OD), 1.94/118.7 (CD_3CN) and 2.50/39.50 (DMSO-d_6) relative to TMS as internal standard wherever applied. Coupling constants are expressed in hertz (Hz) and were directly obtained from the spectra. Splitting patterns are designated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), br (broad) and apparent (app). High-resolution mass spectra (HRMS) were obtained on a quadruple orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 µL/min and spectra were obtained in positive (or negative) ionization mode with a resolution of 15000 (FWHM) using leucine enkephalin as lock mass.

***N*-Carbobenzyloxy-L-alanyl-D,L-2-hydroxyglycine benzyl ester (3.5).**

To a stirred suspension of *N*-carbobenzoxy-L-alanine amide (2.0 g, 9.0 mmol) in CH_2Cl_2 (40 mL) were added benzyl glyoxalate (1.85 g, 11.25 mmol) and a catalytic amount of *p*-toluene sulfonic acid (0.03 g, 0.18 mmol). The reaction mixture was then stirred at room temperature for 7 days and monitored by TLC. Upon completion, the reaction mixture was concentrated under reduced pressure and the resulting crude residue was purified by column chromatography on silica gel (gradient

EtOAc/Hexane, 1:4, v/v; 2:3, v/v; 3:2, v/v) to give **3.5** (2.44 g, 70%) as a white solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.79 and 8.77 (2 x d, J = 8.5 Hz, 1H, N*H*-Gly), 7.44 (d, J = 7.9 Hz, 1H, N*H*-Ala), 7.37-7.33 (m, 10H, 2 x OBn), 6.74-6.65 (m, 1H, α*H*-Gly), 5.53 (br s, 1H, αO*H*-Gly), 5.16-5.14 (m, 2H, CH_2-CO_2Bn), 5.02 (s, 2H, CH_2-Cbz), 4.17-4.07 (m, 1H, α*H*-Ala), 1.19 (app t, J = 7.7 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.5 and 172.5 (CO-Ala), 169.6 and 169.6 (CO-Gly), 155.6 (NHCONH), 137.0 and 137.0 (1C of OCH_2Ph), 135.7 and 135.7 (1C of OCH_2Ph), 128.4 and 128.4 (Ar-C), 128.2 (Ar-C), 128.0 (Ar-C), 127.9 (Ar-C), 127.8 and 127.8 (Ar-C), 127.7 and 127.7 (Ar-C), 71.4 and 71.4 (αC-Gly), 66.1 and 66.1 (CH_2-CO_2Bn), 65.4 and 65.4 (CH_2-Cbz), 49.9 and 49.8 (αC-Ala), 18.1 and 18.0 (CH_3-Ala); HRMS for $C_{20}H_{22}N_2O_6$ $[M-H]^-$ calcd.: 385.1405, found: 385.1402.

3'-*O*-Benzyl-5'-*O*-(*O*-benzyl-*N*,*N*-diisopropyl phosphoramidite)thymidine (3.9).
To a stirred solution of compound **3.8** (300 mg, 0.903 mmol) in anhydrous CH_2Cl_2 (18 mL) was added a solution of benzyloxy-bis(*N*,*N*-diisopropylamino)phosphine (1.68 mL, 0.92 mmol, 0.549 M in anhydrous CH_2Cl_2) followed by a solution of 1H-tetrazole (2.04 mL, 0.92 mmol, 0.45 M in anhydrous AcCN) at 0 °C. The reaction mixture was stirred at room temperature for 4 h. It was then quenched with saturated aq. $NaHCO_3$ and extracted with CH_2Cl_2. The organic layer was dried over Na_2SO_4 and concentrated under reduced pressure. The crude residue was purified by flash chromatography on silica gel (gradient Et_3N/EtOAc/Hexane, 0.2:20:79.8, v/v/v; 0.2:40:59.8, v/v/v; 0.2:80:19.8, v/v/v). The fraction containing the phosphoramidite were combined and concentrated to afford compound **3.9** (497 mg, 72% over two steps) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.67 and 7.56 (2 x s, 1H, H-6), 7.36-7.24 (m, 10H, 2 x OBn), 6.43-6.34 (m, 1H, H-1'), 4.80-4.63 (m, 2H, CH_2P-OBn), 4.58-4.34 (m, 2H, H-5' and H-5''), 4.27 and 4.23 (2 x s, 2H, CH_2-3'-OBn), 3.79-3.78 (m, 1H, H-3'), 3,64-3.60 (m, 3H, H-4' and 2 x *H*-iPr), 2.54-2.38 (m, 1H, H-2'), 2.01-1.95 (m, 1H, H-2''), 1.86 and 1.80 (2 x s, 3H, CH_3-Thy), 1.23-1.15 (m, 12H, CH_3, 4 x iPr); ^{31}P NMR (121 MHz, $CDCl_3$) δ = 148.6 and 148.1; HRMS for $C_{30}H_{40}N_3O_6P$ $[M+H]^+$ calcd.: 570.2727, found: 570.2732.

***N*-Carbobenzyloxy-L-alanyl-D,L-2-acetoxyglycine benzyl ester (3.12).**
Pyridine (21.6 mL) was added dropwise to a stirred suspension of compound **3.5** (2.4 g, 6.21 mmol) in acetic anhydride (30 mL) at 0 °C. The reaction mixture was stirred at the same temperature for 8 h and monitored by TLC. Upon completion, the mixture was concentrated under reduced pressure (bath temp. ~15 °C). The residue was diluted with EtOAc (150 mL) and washed with H_2O (2 x 50 mL), 1N HCl (2 x 50 mL), 5% $NaHCO_3$ (2 x 50 mL) and finally with brine (100 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated *in vacuo*. The resulting crude residue was triturated with cold Et_2O to give **3.12** (2.02 g, 76%) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.59, 7.54 (2 x d, J = 8.5 Hz, 1H, N*H*-Gly), 7.35-7.26 (m, 10H, 2 x OBn), 6.43, 6.39 (2 x d, J = 8.9 Hz, 1H, α*H*-Gly), 5.31 (app t, 1H, N*H*-Ala), 5.20 (s, 2H, CH_2-CO_2Bn), 5.09 (s, 2H, CH_2-Cbz), 4.35-4.30 (m, 1H, α*H*-Ala), 2.05 (s, 3H, CH_3-OAc), 1.38 and 1.36 (2 x d, J = 7.0 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.4 and 172.4 (CO-Ala), 170.3 and 170.3 (CO-OAc), 166.4 (CO-Gly), 156.0 (OCONH), 136.1 (1C of OCH_2Ph), 134.8 (1C of OCH_2Ph), 128.8 (Ar-C), 128.7 (Ar-C), 128.4 (Ar-C), 128.2 (Ar-C), 72.3 and 72.3 (αC-Gly), 68.3 (CH_2-CO_2Bn), 67.4 and 67.4 (CH_2-Cbz), 50.5 (αC-Ala), 20.7 (CH_3-OAc), 18.5 and 18.2 (CH_3-Ala); HRMS for $C_{22}H_{24}N_2O_7$ $[M+Na]^+$ calcd.: 451.1476, found: 451.1474.

5'-*O*-(*N*-Carbobenzyloxy-L-alanyl-D,L-2-aminoglycine benzyl ester)-TMP (3.13) and 3-*N*-(*N*-carbobenzyloxy-L-alanyl-D,L-2-aminoglycine benzyl ester)-TMP triethylammonium salt (3.14).
To a stirred solution of TMP-triethylammonium salt (344 mg, 0.809 mmol) and compound **3.12** (476 mg, 1.496 mmol) in dry DMF (6 mL) was added Et_3N (0.45 mL, 3.234 mmol). The reaction mixture was heated at 40 °C for 24 h. It was then allowed to cool to room temperature, the volatiles were

removed under reduced pressure, and the resulting residue was dissolved in water and lyophilized. The crude product was purified by preparative RP-HPLC (98% H_2O + 2% CH_3CN and 98% CH_3CN + 2% H_2O). The collected eluates were freeze-dried repeatedly until constant mass to afford compound **3.13** (49.6 mg, 8%) as minor product and compound **3.14** (270 mg, 40%) as major product, both as white solids. Data for **3.13**: ^{1}H NMR (600 MHz, CD_3OD) δ = 7.77 and 7.76 (2 x s, 1H, H-6), 7.36-7.26 (m, 10H, 2 x OBn), 6.32-6.29 (m, 1H, H-1'), 6.06 and 6.01 (2 x d, J = 10.0 Hz, 1H, α*H*-Gly), 5.22-5.15 (m, 2H, C*H*$_2$-CO_2Bn), 5.09-5.03 (m, 2H, C*H*$_2$-Cbz), 4.49-4.46 (m, 1H, H-3'), 4.19-4.16 (m, 1H, α*H*-Ala), 4.07-4.02 (m, 2H, H-5' and H-5''), 3.99-3.97 (m, 1H, H-4'), 2.22-2.14 (m, 2H, H-2' and H-2''), 1.90 and 1.89 (2 x s, 3H, C*H*$_3$-Thy), 1.30-1.27 (m, 3H, C*H*$_3$-Ala, merged with Et_3N); ^{13}C NMR (125 MHz, CD_3OD) δ = 175.1 and 175.0 (CO-Ala), 169.2 (CO-Gly), 166.6 and 166.5 (C-4), 158.2 and 158.1 (OCONH), 152.5 (C-2), 138.1 (C-6 and 1C OBn), 136.9 (1C OBn), 129.6 (Ar-C), 129.5 (Ar-C), 129.4 (Ar-C), 129.3 (Ar-C), 129.0 (Ar-C), 128.9 (Ar-C), 112.0 (C-5), 87.5 and 87.4 (d, $^3J_{C,P}$ = 9.4 Hz, C-4'), 86.2 and 86.1 (C-1'), 75.3 (d, $^2J_{C,P}$ = 12.2 Hz, αC-Gly), 72.9 and 72.8 (C-3'), 68.5 (CH_2-CO_2Bn), 67.7 (CH_2-Cbz), 66.6 and 66.5 (d, $^2J_{C,P}$ = 5.4 Hz, C-5'), 51.8 (αC-Ala), 40.9 (C-2'), 18.4 and 18.2 (CH_3-Ala), 12.6 (CH_3-Thy); ^{31}P NMR (202 MHz, CD_3OD) δ = -2.1 and -2.2; HRMS for $C_{30}H_{35}N_4O_{13}P$ $[M-H]^-$ calcd.: 689.1865, found: 689.1879.
Data for **3.14**: ^{1}H NMR (500 MHz, CD_3OD) δ = 7.93 (s, 1H, H-6), 7.37-7.29 (m, 11H, α*H*-Gly and 2 x OBn), 6.34-6.30 (m, 1H, H-1'), 5.28-5.08 (m, 2H, C*H*$_2$-CO_2Bn), 5.06-5.04 (m, 2H, C*H*$_2$-Cbz), 4.50 (br s, 1H, H-3'), 4.26-4.19 (m, 1H, α*H*-Ala), 4.08-4.04 (m, 3H, H-4', H-5' and H-5''), 2.26-2.19 (m, 2H, H-2' and H-2''), 1.95 (s, 3H, C*H*$_3$-Thy), 1.34 and 1.32 (2 x d, J = 7.1 Hz, 3H, C*H*$_3$-Ala, merged with Et_3N); ^{13}C NMR (125 MHz, CD_3OD) δ = 175.4 and 175.3 (CO-Ala), 168.3 and 168.2 (CO-Gly), 164.3 (C-4), 158.3 and 158.2 (OCONH), 151.7 (C-2), 138.1 and 138.0 (1C OBn), 137.5 and 137.4 (C-6), 136.7 (1C OBn), 129.6 (Ar C), 129.5 (Ar C), 129.4 (Ar C), 129.2 (Ar C), 129.1 (Ar C), 129.0 (Ar C), 128.9 (Ar C), 128.8 (Ar C), 111.3 and 111.2 (C-5), 87.9 and 87.8 (d, $^3J_{C,P}$ = 8.8 Hz, C-4'), 87.1 and 87.0 (C-1'), 72.7 and 72.6 (C-3'), 68.9 (CH_2-CO_2Bn), 67.8 and 67.7 (CH_2-Cbz), 65.9 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 58.2 and 58.1 (αC-Gly), 52.0 and 51.6 (αC-Ala), 41.1 and 41.0 (C-2'), 18.0 and 17.9 (CH_3-Ala), 13.0 (CH_3-Thy); ^{31}P NMR (202 MHz, CD_3OD) δ = 0.7; HRMS for $C_{30}H_{35}N_4O_{13}P$ $[M-H]^-$ calcd.: 689.1865, found: 689.1862.

5'-*O*-(L-Alanyl-D,L-2-aminoglycine)-TMP (3.1).
To a stirred solution of **3.13** (30 mg, 0.0434 mmol, 1 eq.) in EtOH/H_2O (5:1, 5 mL), was added 10% Pd/C, Degussa-type (6 mg, 0.2 eq w/w) and evacuation was then carried out with hydrogen atmosphere replacements (3x). The reaction mixture was stirred at room temperature for 4 h under an atmospheric pressure of hydrogen. Upon completion, the catalyst was removed by filtration through a cellulose filter (0.45 μm) and ethanol was removed under reduced pressure (bath temp. ~ 10 °C). The residue was lyophilized to obtain a crude product which was purified by preparative RP-HPLC (98% H_2O + 2% CH_3CN and 98% CH_3CN + 2% H_2O as eluent). The collected eluates were freeze-dried repeatedly until constant mass to afford compound **3.1** as a white solid (8.5 mg, 42%). ^{1}H NMR (300 MHz, D_2O) δ = 7.73 and 7.70 (2 x s, 1H, H-6), 6.37-6.30 (m, 1H, H-1'), 5.73 (d, J = 9.0 Hz, 1H, α*H*-Gly), 4.60-4.54 (m, 1H, H-3'), 4.18-4.02 (m, 4H, H-4', α*H*-Ala, H-5' and H-5''), 2.37-2.33 (m, 2H, H-2' and H-2''), 1.91 (s, 3H, C*H*$_3$-Thy), 1.53 (d, J = 7.1 Hz, 3H, C*H*$_3$-Ala); ^{13}C NMR (125 MHz, D_2O) δ = 172.8 (CO-Gly), 169.9 and 169.8 (CO-Ala), 166.4 (C-4), 151.5 (C-2), 137.1 and 137.0 (C-6), 111.5 and 111.5 (C-5), 85.1 and 85.0 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 84.7 and 84.6 (C-1'), 74.8 (C-3'), 70.9 and 70.8 (αC-Gly), 64.9 and 64.8 (d, $^2J_{C,P}$ = 5.2 Hz, C-5'), 48.8 and 48.7 (αC-Ala), 38.4 and 38.3 (C-2'), 15.9 and 15.8 (CH_3-Ala), 11.4 (CH_3-Thy); ^{31}P NMR (121 MHz, D_2O) δ = -2.0; HRMS for $C_{15}H_{23}N_4O_{11}P$ $[M-H]^-$ calcd.: 465.1028, found: 465.1029.

3-*N*-(L-Alanyl-D,L-2-aminoglycine)-TMP triethylammonium salt (3.15).
Following a similar procedure as the one used for the synthesis of **3.1**, compound **3.15** was obtained as a white solid (179.1 mg, 92%), starting from **3.14** (260 mg, 0.343 mmol), 10% Pd/C, Degussa-type (52 mg, 0.2 eq w/w) in a 5:1 EtOH:H_2O mixture (15 mL). ^{1}H NMR (500 MHz, D_2O) δ = 7.76 and 7.74 (2 x s, 1H, H-6), 6.90 and 6.86 (2 x s, 1H, α*H*-Gly), 6.32 (app t, *J* = 6.8 Hz, 1H, H-1'), 4.55-4.52 (m, 1H, H-3'), 4.21-4.16 (m, 1H, α*H*-Ala), 4.15-4.14 (m, 1H, H-4'), 4.11-4.01 (m, 2H, H-5' and H-5''), 2.36-2.33 (m, 2H, H-2' and H-2''), 1.91 (s, 3H, CH_3-Thy), 1.58 and 1.41 (2 x d, *J* = 7.1 Hz, 3H, CH_3-Ala); ^{13}C NMR (150 MHz, D_2O) δ = 171.0 and 170.9 (CO-Gly), 170.5 and 170.4 (CO-Ala), 164.1 (C-4), 150.7 (C-2), 135.9 (C-6), 110.6 (C-5), 85.6 (C-1'), 85.3 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 70.7 (C-3'), 64.4 (C-5'), 58.3 and 58.2 (αC-Gly), 48.7 and 48.6 (αC-Ala), 38.6 (C-2'), 16.1 and 16.0 (CH_3-Ala), 11.9 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 0.04; HRMS for $C_{15}H_{23}N_4O_{11}P$ $[M-H]^-$ calcd.: 465.1028, found: 465.1025.

***N*-Carbobenzyloxy-L-alanyl-D,L-2-azidoglycine benzyl ester (3.16).**
DPPA (1.56 mL, 7.26 mmol) was added dropwise to a stirred suspension of compound **3.5** (2.0 g, 5.18 mmol) and DBU (1.08 mL, 7.26 mmol) in dry toluene (50 mL) at 0 °C. The reaction mixture was then slowly warmed to room temperature and the stirring was continued for 16 h. The resulting mixture was concentrated under reduced pressure (bath temp. ~15 °C). The residue was diluted with EtOAc (150 mL) and washed with brine (50 mL) and 5% HCl (2 x 50 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated *in vacuo* and the resulting crude residue was purified by column chromatography on silica gel (gradient EtOAc/Hexane, 1:9, v/v; 1:4, v/v; 2:3, v/v) to give **3.16** (1.51 g, 71%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.81-7.77 (2 x d, *J* = 7.1 Hz, 1H, N*H*-Gly), 7.34-7.30 (m, 10H, 2 x OBn), 5.79 and 5.78 (2 x d, *J* = 8.2 Hz, 1H, α*H*-Gly), 5.65 (d, *J* = 6.3 Hz, 1H, N*H*-Ala), 5.21 and 5.20 (2 x s, 2H, CH_2-CO_2Bn), 5.08 and 5.07 (2 x s, 2H, CH_2-Cbz), 4.43-4.38 (m, 1H, α*H*-Ala), 1.36 (app t, *J* = 6.3 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 173.3 (CO-Ala), 166.4 (CO-Gly), 156.1 (OCONH), 136.2 (1C of OCH_2Ph), 134.4 (1C of OCH_2Ph), 128.9 (Ar-C), 128.7 (Ar-C), 128.6 (Ar-C), 128.5 (Ar-C), 128.3 (Ar-C), 128.1 (Ar-C), 68.5 (αC-Gly), 67.3 (CH_2-CO_2Bn), 64.7 and 64.6 (CH_2-Cbz), 50.4 (αC-Ala), 18.6, 18.3 (CH_3-Ala); HRMS for $C_{20}H_{21}N_5O_5$ $[M+Na]^+$ calcd.: 434.1435, found: 434.1432.

***N*-Carbobenzyloxy-L-alanyl-D,L-2-aminoglycine benzyl ester (3.17).**
Triphenylphospine (0.92 g, 3.50 mmol) was added to a stirred solution of **3.16** (1.2 g, 2.92 mmol) and water (0.26 mL, 14.6 mmol) in THF (15 mL). The reaction mixture was refluxed for 2 h, then cooled to room temperature and concentrated under reduced pressure. Water (100 mL) was added to the residue and the solution was extracted with CH_2Cl_2 (3 x 100 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated *in vacuo* and the resulting crude residue was purified by column chromatography on silica gel (gradient MeOH/CH_2Cl_2, 1:99, v/v; 1:49, v/v; 1:19, v/v) to give **3.17** (0.79 g, 70%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.77 and 7.73 (2 x d, *J* = 8.3 Hz, 1H, N*H*-Gly), 7.31-7.29 (m, 10H, 2 x OBn), 5.74 (app t, *J* = 9.0 Hz, 1H, N*H*-Ala), 5.17- 5.07 (m, 3H, α*H*-Gly and CH_2-CO_2Bn), 5.06 and 5.04 (2 x s, 2H, CH_2-Cbz), 4.30-4.20 (m, 1H, α*H*-Ala), 2.84 (br s, 2H, αNH_2-Gly), 1.31 and 1.30 (2 x d, *J* = 7.0 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.7 (CO-Ala), 170.5 (CO-Gly), 156.0 (OCONH), 136.3 (1C of OCH_2Ph), 135.1 (1C of OCH_2Ph), 128.6 (Ar-C), 128.5 (Ar-C), 128.3 (Ar-C), 128.2 (Ar-C), 128.1 (Ar-C), 67.6 (αC-Gly), 67.0 (CH_2-CO_2Bn), 60.3 and 60.2 (CH_2-Cbz), 50.5 (αC-Ala), 18.5 (CH_3-Ala); HRMS for $C_{20}H_{23}N_3O_5$ $[M+Na]^+$ calcd.: 408.1530, found: 408.1523.

Thymidine-5'-*O*-(*N*-carbobenzyloxy-L-alanyl-R-2-aminoglycine benzyl ester) phosphoramidate (3.18a) and thymidine-5'-*O*-(*N*-carbobenzyloxy-L-alanyl-S-2-aminoglycine benzyl ester) phosphoramidate (3.18b).

o a stirred suspension of TMP-triehtylammonium salt (400 mg, 0.737 mmol) and compound **3.17** (640 mg, 1.659 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (15 mL), triethylamine (0.1 mL, 0.737 mmol) was added to facilitate dissolution, followed by DCC (532 mg, 2.579 mmol). The reaction mixture was heated at 85 ºC for 2.5 h. The reaction progress was monitored by TLC (CH_2Cl_2:MeOH:H_2O 17:7:1). Upon completion, the mixture was cooled to room temperature and the solvent was removed by rotary evaporation. The resulting residue was resuspended in water and lyophilized to give a crude material which was purified by column chromatography on silica gel using the following gradient (CH_2Cl_2:MeOH:H_2O 19:1:0, v/v/v; 9:1:0, v/v/v; 17:7:1, v/v/v) to provide the desired nucleoside phosphoramidates as salts. Semi-preparative RP-HPLC (98% H_2O + 2% CH_3CN and 98% CH_3CN + 2% H_2O as eluent) was employed for further separation, to obtain the pure diastereomers as a ~1:1 ratio, as white powders (131 mg of compound **3.18a**, 129 mg of compound **3.18b**, overall yield **3.18a** + **3.18b** 260 mg, 51%). The isolated products were freeze-dried repeatedly until constant mass. Data for **3.18a**: ^{1}H NMR (500 MHz, CD_3CN + D_2O, 50 ºC) δ = 7.55 (s, 1H, H-6), 7.32-7.28 (m, 10H, 2 x OBn), 6.14 (app t, J = 6.8 Hz, 1H, H-1'), 5.21 and 5.16 (2 x d, J = 10.9 Hz, 1H, α*H*-Gly), 5.14- 4.99 (m, 4H, C*H*$_2$-CO_2Bn and C*H*$_2$-Cbz), 4.38-4.36 (m, 1H, H-3'), 4.11-4.06 (m, 1H, α*H*-Ala), 3.93-3.91 (m, 1H, H-4'), 3.87-3.82 (m, 2H, H-5' and H-5''), 2.17-2.13 (m, 2H, H-2' and H-2''), 1.80 and 1.79 (2 x s, 3H, C*H*$_3$-Thy), 1.20 (app t, J = 7.0 Hz, 3H, C*H*$_3$-Ala); ^{13}C NMR (150 MHz, CD_3CN + D_2O, 10 ºC) δ = 174.6 (CO-Ala), 171.3 (CO-Gly), 166.3 (C-4), 157.5 (OCONH), 151.5 (C-2), 137.9 (C-6), 137.5 (1C of OCH_2Ph), 136.2 (1C of OCH_2Ph), 129.5 (Ar-C), 129.4 (Ar-C), 129.2 (Ar-C), 129.0 (Ar-C), 128.7 (Ar-C), 111.9 (C-5), 86.4 (d, $^3J_{C,P}$ = 8.9 Hz, C-4'), 84.8 (C-1'), 71.7 (C-3'), 68.3 (CH_2-CO_2Bn), 67.4 (CH_2-Cbz), 64.9 (d, $^2J_{C,P}$ = 3.9 Hz, C-5'), 61.1 (d, $^2J_{C,P}$ = 9.8 Hz, αC-Gly), 51.1 (αC-Ala), 39.8 (C-2'), 18.1 (CH_3-Ala), 12.6 (CH_3-Thy); ^{31}P NMR (202 MHz, CD_3CN + D_2O, 50 ºC) δ = 4.4; $[\alpha]^{20}_{589}$ = -0.147° (c = 1, CH_3OH); HRMS for $C_{30}H_{36}N_5O_{12}P$ $[M-H]^-$ calcd.: 688.2025, found: 688.2040.

Data for **3.18b**: ^{1}H NMR (300 MHz, CD_3CN + D_2O, 25 ºC) δ = 7.54 (s, 1H, H-6), 7.32-7.27 (m, 10H, 2 x OBn), 6.14 (app t, J = 6.8 Hz, 1H, H-1'), 5.21 and 5.16 (2 x d, J = 10.9 Hz, 1H, α*H*-Gly), 5.11-4.93 (m, 4H, C*H*$_2$-CO_2Bn and C*H*$_2$-Cbz), 4.37-4.32 (m, 1H, H-3'), 4.11-4.01 (m, 1H, α*H*-Ala), 3.95-3.90 (m, 1H, H-4'), 3.89-3.85 (m, 2H, H-5' and H-5''), 2.17-2.11 (m, 2H, H-2' and H-2''), 1.79 and 1.77 (2 x s, 3H, C*H*$_3$-Thy), 1.19 (app t, J = 7.0 Hz, 3H, C*H*$_3$-Ala); ^{13}C NMR (150 MHz, CD_3CN + D_2O, 10 ºC) δ = 174.7 (CO-Ala), 171.3 (CO-Gly), 166.3 (C-4), 157.5 (OCONH), 152.0 (C-2), 137.9 (C-6), 137.5 (1C of OCH_2Ph), 136.3 (1C of OCH_2Ph), 129.5 (Ar-C), 129.4 (Ar-C), 129.0 (Ar-C), 128.7 (Ar-C), 112.0 (C-5), 86.3 (d, $^3J_{C,P}$ = 8.7 Hz, C-4'), 85.5 (C-1'), 71.6 (C-3'), 68.4 (CH_2-CO_2Bn), 67.4 (CH_2-Cbz), 64.9 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 61.1 (d, $^2J_{C,P}$ = 9.1 Hz, αC-Gly), 51.1 (αC-Ala), 39.7 (C-2'), 18.1 (CH_3-Ala), 12.6 (CH_3-Thy); ^{31}P NMR (121 MHz, CD_3CN + D_2O, 25 ºC) δ = 3.2; $[\alpha]^{20}_{589}$ = -0.052° (c = 1, CH_3OH); HRMS for $C_{30}H_{36}N_5O_{12}P$ $[M-H]^-$ calcd.: 688.2025, found: 688.2019.

Thymidine-5'-*O*-(L-alanyl-R-2-aminoglycine) phosphoramidate (3.2a).

Following a similar procedure as the one used for the synthesis of **3.1**, compound **3.2a** was obtained as a white solid (45.5 mg, 54%), starting from **3.18a** (125 mg, 0.136 mmol, 1 eq.), 10% Pd/C, Degussa-type (25 mg, 0.2 eq w/w) in a 5:1 EtOH/H_2O mixture (10 mL). ^{1}H NMR (300 MHz, D_2O) δ = 7.63 (d, J = 1.1 Hz, 1H, H-6), 6.14 (app t, J = 6.8 Hz, 1H, H-1'), 5.23 (d, J = 11.3 Hz, 1H, α*H*-Gly), 4.50-4.45 (m, 1H, H-3'), 4.06-4.02 (m, 1H, H-4'), 4.00-3.88 (m, 3H, α*H*-Ala, H-5' and H-5''), 2.31-2.26 (m, 2H, H-2' and H-2''), 1.82 (d, J = 1.1 Hz, 3H, C*H*$_3$-Thy), 1.44 (d, J = 7.1 Hz, 3H, C*H*$_3$-Ala); ^{13}C NMR (150 MHz, D_2O) δ = 172.0 (d, $^3J_{C,P}$ = 9.1 Hz, CO-Gly), 169.9 (CO-Ala), 166.3 (C-4), 151.5 (C-2), 137.3 (C-6), 111.5 (C-5), 85.2 (d, $^3J_{C,P}$ = 9.0 Hz, C-4'), 84.8 (C-1'), 70.8 (C-3'), 64.2 (d, $^2J_{C,P}$ =

4.6 Hz, C-5'), 61.1 (d, $^{2}J_{C,P}$ = 6.5 Hz, αC-Gly), 48.7 (αC-Ala), 38.4 (C-2'), 16.0 (CH_3-Ala), 11.5 (CH_3-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 4.4; HRMS for $C_{15}H_{24}N_5O_{10}P$ $[M-H]^-$ calcd.: 464.1188, found: 464.1185.

Thymidine-5'-*O*-(L-alanyl-S-2-aminoglycine) phosphoramidate (3.2b).
Following a similar procedure as the one used for the synthesis of **3.1**, compound **3.2b** was obtained as a white solid (43.0 mg, 51%), starting from **3.18b** (125 mg, 0.136 mmol, 1 eq.), 10% Pd/C, Degussa-type (25 mg, 0.2 eq w/w) in a 5:1 EtOH/H_2O mixture (10 mL). ^{1}H NMR (300 MHz, D_2O) δ = 7.68 (s, 1H, H-6), 6.25 (app t, *J* = 6.9 Hz, 1H, H-1'), 5.24 and 5.19 (2 x d, *J* = 11.0 Hz, 1H, α*H*-Gly), 4.52-4.47 (m, 1H, H-3'), 4.11-4.07 (m, 1H, H-4'), 4.06-3.90 (m, 3H, α*H*-Ala, H-5' and H-5''), 2.34-2.29 (m, 2H, H-2' and H-2''), 1.86 (s, 3H, CH_3-Thy), 1.47 and 1.46 (2 x d, *J* = 7.0 Hz, 3H, CH_3-Ala); ^{13}C NMR (150 MHz, D_2O) δ = 172.2 (d, $^{3}J_{C,P}$ = 9.2 Hz, CO-Gly), 169.9 (CO-Ala), 166.4 (C-4), 151.6 (C-2), 137.3 (C-6), 111.5 (C-5), 85.3 (d, $^{3}J_{C,P}$ = 8.4 Hz, C-4'), 84.9 (C-1'), 71.0 (C-3'), 64.3 (d, $^{2}J_{C,P}$ = 4.7 Hz, C-5'), 61.1 (d, $^{2}J_{C,P}$ = 9.1 Hz, αC-Gly), 48.7 (αC-Ala), 38.5 (C-2'), 16.0 (CH_3-Ala), 11.5 (CH_3-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 4.2; HRMS for $C_{15}H_{24}N_5O_{10}P$ $[M-H]^-$ calcd.: 464.1188, found: 464.1187.

***N*-Boc-L-Alanyl-D,L-2-acetoxyglycine methyl ester (3.19).**
To a stirred solution of Boc-L-Ala-L-Ser-OMe (3.9 g, 13.43 mmol) in dry ethyl acetate (220 mL) were added molecular sieves 4 Å (5 g) followed by $Pb(OAc)_4$ (17.89 g, 40.30 mmol) under an inert atmosphere. The reaction mixture was heated at reflux for 2 h, and cooled to room temperature. The solid was filtered off through a pad of Celite and the organic layer was stirred with 10% aq. citric acid until the organic layer became nearly colourless. The organic layer was separated, washed with 10% aq. citric acid, water and brine. It was then dried over Na_2SO_4, filtered and concentrated to give a crude diastereomeric mixture of **3.19** (4.27 g, quant.) as a colourless foam, which was used in the next step without any further purification. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.70-7.52 (m, 1H, N*H*-Gly), 6.40 and 6.38 (2 x d, *J* = 9.1 Hz, 1H, α*H*-Gly), 5.31 (app t, *J* = 6.9 Hz, 1H, N*H*-Ala), 4.27-4.13 (m, 1H, α*H*-Ala), 3.81 and 3.80 (2 x s, 3H, OCH_3-Gly), 2.11 (s, 3H, CH_3-OAc), 1.45 (s, 9H, ^{t}Bu), 1.38 and 1.37 (2 x d, *J* = 7.1 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.8 (CO-Ala), 170.3 and 170.2 (CO-Gly), 167.2 and 167.1 (CO-OAc), 155.6 (OCONH), 88.7 (1C-^{t}Bu), 72.3 and 72.2 (αC-Gly), 53.4 and 53.3 (OCH_3-Gly), 50.3 and 50.1 (αC-Ala), 28.4 (^{t}Bu), 20.7 (CH_3-OAc), 17.8 and 17.6 (CH_3-Ala); HRMS for $C_{13}H_{22}N_2O_7$ $[M+Na]^+$ calcd.: 341.1319, found: 341.1318.

***N*-Boc-L-Alanyl-D,L-2-*p*-anisylaminoglycine methyl ester (3.20).**
DIPEA (3.01 mL, 17.28 mmol) was added to a stirred solution of **3.19** (2.2 g, 6.91 mmol) and 4-methoxybenzylamine (1.0 mL, 7.60 mmol) in dry DMF (20 mL) at room temperature. The reaction mixture was heated at 40 °C for 24 h and was then concentrated under reduced pressure. Water (200 mL) was added to the residue and extracted with ethyl acetate (3 x 120 mL). The combined organic layers were washed with brine, dried over Na_2SO_4, filtered and concentrated and the resulting crude was purified by column chromatography on silica gel (gradient EtOAc/Hexane, 2:3, v/v; 3:2, v/v; 4:1, v/v) to give a diastereomeric mixture of **3.20** (2.07 g, 76%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.22 (d, *J* = 8.5 Hz, 1H, *o*-H PMB), 7.04- 6.93 (m, 1H, N*H*-Gly), 6.84 (d, *J* = 8.5 Hz, 1H, *m*-H PMB), 5.27 (d, *J* = 7.8 Hz, 1H, α*H*-Gly), 5.08-4.99 (m, 1H, N*H*-Ala), 4.20-4.18 (m, 1H, α*H*-Ala), 3.78 (s, 3H, OCH_3-PMB), 3.73-3.70 (m, 5H, OCH_3-Gly and CH_2-PMB), 2.37 (br s, 1H, N*H*-PMB), 1.46 and 1.45 (2 x s, 9H, ^{t}Bu), 1.37 (app t, *J* = 7.0 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 173.1 (CO-Ala), 170.6 and 170.5 (CO-Gly), 159.0 (Ar-C), 155.6 (OCONH), 131.3 and 131.2 (1C of PMB), 129.7 (Ar-C), 129.3 (Ar-C), 114.3 (Ar-C), 113.9 (Ar-C), 80.4 (1C-^{t}Bu), 64.4 and

64.3 (αC-Gly), 55.4 (OCH_3-PMB), 52.9 (OCH_3-Gly), 50.3 (αC-Ala), 48.5 (CH_2-PMB), 28.4 (^{t}Bu), 18.3 and 18.1 (CH_3-Ala); HRMS for $C_{19}H_{29}N_3O_6$ $[M-H]^-$ calcd.: 394.1983, found: 394.1973.

_N_-Boc-L-Alanyl-D,L-2-aminoglycine methyl ester (3.21).
To a stirring solution of **3.20** (1.8 g, 2.44 mmol) in EtOH (80 mL) was added 10% Pd/C (0.18 g, 0.1 eq w/w) and the mixture was hydrogenated at atmospheric pressure using a balloon filled with H_2 for 5 h at room temperature. The catalyst was then removed by filtration through a pad of Celite and the filtrate was concentrated under reduced pressure. The resulting crude residue was purified by column chromatography on silica gel (gradient MeOH/CH_2Cl_2, 1:99, v/v; 1:49, v/v; 1:24, v/v) to give a diastereomeric mixture of **3.21** (1.2 g, 96%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.41 (br s, 1H, N*H*-Gly), 5.28-5.24 (m, 2H, α*H*-Gly and N*H*-Ala), 4.20-4.18 (m, 1H, α*H*-Ala), 3.79 (s, 3H, OCH_3-Gly), 2.24 (br s, 1H, αNH_2-Gly), 1.45 (s, 9H, ^{t}Bu), 1.36 (d, J = 7.1 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 173.1 and 173.0 (CO-Ala), 171.2 and 171.1 (CO-Gly), 155.5 (OCONH), 80.3 (1C-^{t}Bu), 60.2 and 60.1 (αC-Gly), 53.0 (OCH_3-Gly), 50.1 (αC-Ala), 28.3 (^{t}Bu), 18.3 and 18.2 (CH_3-Ala); HRMS for $C_{19}H_{29}N_3O_6$ $[M+Na]^+$ calcd.: 298.1377, found: 298.1390.

Thymidine-5'-*O*-(*N*-Boc-L-alanyl-D,L-2-aminoglycine methyl ester) phosphoramidate triethylammonium salt (3.22).
Following a similar procedure as the one used for the synthesis of **3.18a-b**, the triethylammonium salt of compound **3.22** was obtained as a diastereomeric mixture (538 mg, 74%, white solid), starting from TMP-triehtylammonium salt (560 mg, 1.07 mmol), compound **3.21** (647 mg, 2.35 mmol), triethylamine (163 µL, 1.17 mmol) and DCC (881 mg, 4.27 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (15 mL). ^{1}H NMR (500 MHz, CD_3OD) δ = 7.68 (d, J = 1.1 Hz, 1H, H-6), 6.35-6.31 (m, 1H, H-1'), 5.30 and 5.22 (2 x d, J = 11.1 Hz, 1H, α*H*-Gly) 4.48-4.45 (m, 1H, H-3'), 4.10-4.05 (m, 1H, α*H*-Ala), 4.02-4.00 (m, 1H, H-4'), 3.99-3.95 (m, 2H, H-5' and H-5''), 3.71 and 3.69 (2 x s, 3H, OCH_3-Gly), 2.29-2.17 (m, 2H, H-2' and H-2''), 1.94 (d, J = 1.1 Hz, 3H, CH_3-Thy), 1.44 and 1.43 (2 x s, 9H, ^{t}Bu), 1.29 (app t, J = 7.1 Hz, 3H, CH_3-Ala, merged with Et_3N); ^{13}C NMR (125 MHz, CD_3OD) δ = 175.2 (CO-Ala), 172.1 and 172.0 (CO-Gly), 166.5 (C-4), 157.5 (OCONH), 152.5 (C-2), 138.3 (C-6), 111.9 (C-5), 87.6 (app t, $^3J_{C,P}$ = 9.5 Hz, C-4'), 86.2 and 86.0 (C-1'), 80.6 (1C-^{t}Bu), 73.0 and 72.9 (C-3'), 65.7 (app t, $^2J_{C,P}$ = 6.9 Hz, C-5'), 61.7 (αC-Gly), 53.1 and 53.0 (OCH_3-Gly), 51.3 (αC-Ala), 40.8 (C-2'), 28.7 (^{t}Bu), 18.4 (CH_3-Ala), 12.7 (CH_3-Thy); ^{31}P NMR (202 MHz, CD_3OD) δ = 3.4; HRMS for $C_{21}H_{34}N_5O_{12}P$ $[M-H]^-$ calcd.: 578.1869, found: 578.1873.

Thymidine-5'-(L-alanyl-D,L-2-aminoglycine methyl ester) phosphoramidate TFA salt (3.23). To a stirred solution of **3.22** (220 mg, 0.323 mmol) and thioanisole (42 µL, 0.356 mmol) in CH_2Cl_2 (4.5 mL), was added TFA (1.5 mL) at 0 °C. The reaction mixture was stirred at 0 °C for 1 h and then at room temperature for 3 h. After completion of the reaction, all the volatiles were removed under reduced pressure and coevaporated three times with toluene to remove residual TFA. The residue was redissolved in H_2O and lyophilized (2x) to obtain a crude residue which was purified by preparative RP-HPLC (98% H_2O + 2% CH_3CN and 98% CH_3CN + 2% H_2O). The collected eluates were freeze-dried repeatedly until constant mass to obtain a diastereomeric mixture of **3.23** (158 mg, 85%, white solid). ^{1}H NMR (600 MHz, D_2O) δ = 7.72 (s, 1H, H-6), 6.29 (app t, J = 7.0 Hz, 1H, H-1'), 5.64 (d, J = 9.5 Hz, 1H, α*H*-Gly), 4.54-4.52 (m, 1H, H-3'), 4.18-4.15 (m, 1H, α*H*-Ala), 4.14-4.12 (m, 1H, H-4'), 4.07-4.03 (m, 2H, H-5' and H-5''), 3.84 (s, 3H, OCH_3-Gly), 2.32-2.30 (m, 2H, H-2' and H-2''), 1.86 (s, 3H, CH_3-Thy), 1.53 and 1.52 (2 x d, J = 7.2 Hz, 3H, CH_3-Ala); ^{13}C NMR (150 MHz, D_2O) δ = 171.6 and 171.4 (CO-Ala), 166.3 (CO-Gly), 166.2 (C-4), 151.5 (C-2), 137.1 (C-6), 111.4 (C-5), 85.2 (app t, $^3J_{C,P}$ = 9.0 Hz, C-4'), 84.8 (C-1'), 70.9 (C-3'), 64.7 (d, $^2J_{C,P}$ = 4.5 Hz, C-5'), 56.8 (d, $^2J_{C,P}$ = 7.5 Hz, αC-Gly), 54.1 (OCH_3-Gly), 48.5 (αC-Ala), 38.6 (C-2'), 15.9 and 15.8 (CH_3-Ala), 11.4 (CH_3-Thy);

^{31}P NMR (202 MHz, D_2O) δ = -0.1; HRMS for $C_{16}H_{26}N_5O_{10}P$ $[M-H]^-$ calcd.: 478.1344, found: 478.1346.

_N_-Carbobenzyloxy-L-alanyl-D,L-2-mercaptoethanolglycine benzyl ester (3.24).
Et_3N (0.55 mL, 3.92 mmol) was added dropwise to a stirred solution of **3.12** (1.4 g, 3.27 mmol) and 2-mercaptoethanol (0.25 mL, 3.59 mmol) in dry DMF (12 mL) at 0 °C. The reaction mixture was then slowly warmed to room temperature and the stirring was continued for 16 h. The resulting mixture was concentrated under reduced pressure. Water (100 mL) was added to the residue and extracted with ethyl acetate (3 x 100 mL). The combined organic layers were washed with brine, dried over Na_2SO_4, filtered and concentrated and the resulting crude was purified by column chromatography on silica gel (gradient EtOAc/Hexane, 1:4, v/v; 2:3, v/v; 3:2, v/v) to give a diastereomeric mixture of **3.24** (1.2 g, 70%) as a colorless foam. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.83 and 8.79 (2 x d, J = 8.2 Hz, 1H, N*H*-Gly), 7.47 (d, J = 7.3 Hz, 1H, N*H*-Ala), 7.39-7.32 (m, 10H, 2 x OBn), 5.46 and 5.44 (2 x d, J = 10.0 Hz, 1H, α*H*-Gly), 5.19- 5.16 (m, 2H, CH_2-CO_2Bn), 5.00 (s, 2H, CH_2-Cbz), 4.96-4.93 (m, 1H, O*H*), 4.19-4.12 (m, 1H, α*H*-Ala), 3.54-3.48 (m, 2H, OC$H_2$$CH_2$S), 2.72-2.65 (m, 2H, OCH_2CH_2S), 1.19 (app t, J = 7.8 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.7 and 172.6 (CO-Ala), 168.5 and 168.4 (CO-Gly), 155.7 and 155.6 (OCONH), 137.0 (1C of OCH_2Ph), 135.6 and 135.5 (1C of OCH_2Ph), 128.5 (Ar-C), 128.4 (Ar-C), 128.2 (Ar-C), 127.9 (Ar-C), 127.7 (Ar-C), 66.8 and 66.7 (CH_2-CO_2Bn), 65.5 (CH_2-Cbz), 60.4 and 60.3 (O*C*$H_2$$CH_2$S), 53.2 and 53.1 (αC-Gly), 49.9 and 49.8 (αC-Ala), 32.7 and 32.6 (OCH_2*C*H_2S), 18.1 and 18.0 (CH_3-Ala); HRMS for $C_{22}H_{26}N_2O_6S$ $[M-H]^-$ calcd.: 445.1439, found: 445.1443.

3'-*O*-Benzyl-5'-*O*-(*N*-carbobenzyloxy-L-alanyl-D,L-2-mercaptoethanolglycine benzyl ester) TMP triethylammonium salt (3.25).
$POCl_3$ (169 µL, 3.92 mmol) was added dropwise to a stirred solution of **3.8** (0.5 g, 1.50 mmol) in dry trimethyl phosphate (5 mL) at -20 °C. The reaction mixture was then slowly warmed to room temperature and the stirring was continued for 8 h until reaction completion (monitored by TLC). The solution was again cooled to 0 °C and a mixture of compound **3.24** (0.672 g, 1.50 mmol) and *N*-methyl imidazole (0.36 mL, 4.51 mmol) dissolved in dry CH_2Cl_2 (14 mL) was then added. The resulting mixture was stirred at room temperature for 16 h. It was then quenched with water at 0 °C and the volatiles were removed under reduced pressure, water was added to the residue which was then lyophilized (3x). The resulting crude residue was purified first by column chromatography on silica gel (gradient MeOH/CH_2Cl_2, 1:99, v/v; 3:97, v/v; 8:92, v/v), then by preparative RP-HPLC (25 mmol TEAB in 98% H_2O + 2% CH_3CN and 25 mmol TEAB in 98% CH_3CN + 2% H_2O as eluent). The collected eluates were freeze-dried repeatedly until constant mass, to obtain a diastereomeric mixture of **3.25** (935 mg, 66%) as a white solid. ^{1}H NMR (300 MHz, CD_3OD) δ = 7.76 (s, 1H, H-6), 7.35-7.25 (m, 15H, 3 x OBn), 6.45 (dd, J = 8.7, 5.8 Hz, 1H, H-1'), 5.51 (d, J = 13.4 Hz, 1H, α*H*-Gly), 5.24- 5.13 (m, 2H, CH_2-CO_2Bn), 5.06 (s, 2H, CH_2-Cbz), 4.57-4.55 (m, 2H, CH_2-3'OBn), 4.34-4.31 (m, 1H, α*H*-Ala), 4.25-4.18 (m, 1H, H-3' and H-4'), 4.07-4.01 (m, 2H, H-5' and H-5''), 4.00-3.90 (m, 2H, OC$H_2$$CH_2$S), 2.93-2.78 (m, 2H, OCH_2CH_2S), 2.40-2.17 (m, 2H, H-2' and H-2''), 1.91 (s, 3H, CH_3-Thy), 1.34 and 1.32 (2 x d, J = 7.2 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, CD_3OD) δ = 175.2 (CO-Ala), 169.7 and 169.6 (CO-Gly), 166.4 and 166.3 (C-4), 158.1 (OCONH), 152.5 and 152.4 (C-2), 139.4 (1C Ar-C), 138.2 (1C Ar-C), 137.9 (C-6), 136.8 (1C Ar-C), 129.6 (Ar-C), 129.5 (Ar-C), 129.4 (Ar-C), 129.3 (Ar-C), 129.2 (Ar-C), 129.0 (Ar-C), 128.9 (Ar-C), 128.8 (Ar-C), 128.7 (Ar-C), 112.1 and 112.0 (C-5), 86.2 and 86.1 (C-1'), 85.1 (d, $^3J_{C,P}$ = 8.7 Hz, C-4'), 81.1 and 81.0 (C-3'), 72.2 and 72.1 (CH_2-3'OBn), 68.6 and 68.5 (CH_2-CO_2Bn), 67.6 (CH_2-Cbz), 66.9 (d, $^2J_{C,P}$ = 5.5 Hz, C-5'), 65.8 and 65.7 (d, $^2J_{C,P}$ = 5.4 Hz, O$CH_2$$CH_2$S), 54.8 and 54.7 (αC-Gly), 52.0 and 51.9 (αC-Ala), 38.1

(C-2'), 32.1 and 32.0 (OCH_2CH_2S), 18.3 and 18.2 (CH_3-Ala), 12.7 (s, CH_3-Thy); ^{31}P NMR (121 MHz, CD_3OD) δ = -0.5; HRMS for $C_{39}H_{45}N_4O_{13}PS$ $[M-H]^-$ calcd.: 839.2368, found: 839.2376.

_N_-Boc-L-Alanyl-D,L-2-mercaptoethanolglycine methyl ester (3.26).
Following a similar procedure as the one used for the synthesis of **3.24**, a diastereomeric mixture of compound **3.26** was obtained as a colorless foam (3.5 g, 77%), starting from **3.19** (4.28 g, 13.43 mmol), 2-mercaptoethanol (1.04 mL, 14.77 mmol) and Et_3N (2.24 mL, 16.12 mmol) in dry DMF (12 mL). ^1H NMR (300 MHz, $CDCl_3$) δ = 7.63 (br s, 1H, N*H*-Gly), 5.56- 5.50 (m, 1H, α*H*-Gly), 5.10 and 5.06 (2 x d, *J* = 7.0 Hz, 1H, N*H*-Ala), 4.28-4.13 (m, 1H, α*H*-Ala), 3.92-3.77 (m, 5H, OCH_2CH_2S and OCH_3-Gly), 2.90-2.85 (m, 2H, OCH_2CH_2S), 1.46 and 1.45 (2 x s, ^{t}Bu), 1.38 and 1.37 (2 x d, *J* = 7.0 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.5 (CO-Ala), 169.8 and 169.6 (CO-Gly), 155.6 (OCONH), 80.7 (1C-^{t}Bu), 62.1 (OCH_2CH_2S), 53.2 (OCH_3-Gly), 52.9 (αC-Gly), 50.4 (αC-Ala), 34.2 and 34.1 (OCH_2CH_2S), 28.4 (^{t}Bu), 18.0 and 17.9 (CH_3-Ala); HRMS for $C_{13}H_{24}N_2O_6S$ $[M+Na]^+$ calcd.: 359.1247, found: 359.1240.

3'-*O*-Benzoyl-5'-*O*-(*N*-Boc-L-alanyl-D,L-2-mercaptoethanolglycine methyl ester)TMP (3.29).
Following a similar procedure as the one used for the synthesis of **3.25**, a diastereomeric mixture of compound **3.29** was obtained as a white solid (101.5 mg, 12%), starting from **3.27** (394 mg, 1.138 mmol), $POCl_3$ (128 µL, 1.366 mmol) in dry trimethyl phosphate (4 mL) and compound **3.26** (383 mg, 1.138 mmol), *N*-methyl imidazole (272 µL, 1.366 mmol) in dry CH_2Cl_2 (10 mL). The resulting crude residue was purified first by column chromatography on silica gel (gradient MeOH/CH_2Cl_2, 1:99, v/v; 3:97, v/v; 8:92, v/v), then by preparative RP-HPLC (98% H_2O + 2% CH_3CN and 98% CH_3CN + 2% H_2O as eluent). The collected eluates were freeze-dried repeatedly until constant mass. ^1H NMR (300 MHz, CD_3OD) δ = 8.06 (d, *J* = 7.5 Hz, 2H, *oH*-OBz), 7.78 (s, 1H, H-6), 7.64 (t, *J* = 7.4 Hz, 1H, *pH*-OBz), 7.64 (t, *J* = 7.6 Hz, 2H, *mH*-OBz), 6.45 (app t, *J* = 7.5 Hz, 1H, H-1'), 5.67-5.62 (m, 1H, H-3'), 5.59-5-56 (m, 1H, α*H*-Gly), 4.67-4.37 (m, 1H, H-4'), 4.35-4.23 (m, 2H, H-5' and H-5''), 4.18-4.04 (m, 3H, OCH_2CH_2S and α*H*-Ala), 3.74 (s, 3H, OCH_3-Gly), 3.02-2.92 (m, 2H, OCH_2CH_2S), 2.58-2.50 (m, 2H, H-2' and H-2''), 1.95 (s, 3H, CH_3-Thy), 1.42 (s, 9H, ^{t}Bu), 1.30 and 1.29 (2 x d, *J* = 7.0 Hz, 3H, CH_3-Ala); ^{13}C NMR (150 MHz, D_2O) δ = 175.5 and 175.4 (CO-Ala), 170.4 and 170.3 (CO-Gly), 167.3 (CO-OBz), 166.4 and 166.3 (C-4), 157.6 (OCONH), 152.6 and 152.5 (C-2), 137.8 (C-6), 134.6 and 134.5 (1C Ar-C), 131.0 (Ar-C), 130.9 (Ar-C), 130.7 (Ar-C), 130.6 (Ar-C), 129.7 (Ar-C), 129.6 (Ar-C), 112.4 and 112.3 (C-5), 86.1 and 86.0 (C-1'), 85.1 (d, $^3J_{C,P}$ = 7.9 Hz, C-4'), 80.6 (1C-^{t}Bu), 77.6 and 77.4 (C-3'), 66.9 (d, $^2J_{C,P}$ = 3.0 Hz, C-5'), 66.1 (d, $^2J_{C,P}$ = 6.2 Hz, OCH_2CH_2S), 54.7 and 54.6 (αC-Gly), 53.4 and 53.3 (OCH_3-Gly), 51.5 (αC-Ala), 38.3 and 38.2 (C-2'), 31.9 (OCH_2CH_2S), 28.7 (^{t}Bu), 18.2 and 18.1 (CH_3-Ala), 12.7 and 12.6 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = -0.5; HRMS for $C_{30}H_{41}N_4O_{14}PS$ $[M-H]^-$ calcd.: 743.2005, found: 743.2014.

5'-*O*-(L-Alanyl-D,L-2-mercaptoethanolglycine) TMP (3.4).
Following a similar procedure as the one used for the synthesis of **3.23**, the TFA salt of the crude diastereomeric amine **3.30** was obtained as a sticky mass (~100 mg, quant.), starting from compound **3.29** (100 mg, 0.134 mmol), thioanisole (16 µL, 0.134 mmol) and TFA (0.5 mL) in CH_2Cl_2 (1.5 mL). HRMS for $C_{25}H_{33}N_4O_{12}PS$ $[M-H]^-$ calcd.: 643.1480, found: 643.1472.
The so obtained residue **3.30** was treated without further purification with LiOH (19.4 mg, 0.809 mmol) in a 1:1 MeOH:H_2O mixture (2 mL) at 0 °C and then the solution was stirred at room temperature for 3 h. After completion of the reaction, the mixture was neutralized with 5% acetic acid and evaporated to dryness *in vacuo*. The crude residue was purified by preparative RP-HPLC (98% H_2O + 2% CH_3CN and 98% CH_3CN + 2% H_2O). The collected eluates were freeze-dried repeatedly until constant mass to afford a diastereomeric mixture of compound **3.4** (51 mg, 72% over two steps)

as a white solid. ^{1}H NMR (500 MHz, D_2O) δ = 7.60 (s, 1H, H-6), 6.39 (app t, J = 6.9 Hz, 1H, H-1'), 5.22 (d, J = 20.5 Hz, 1H, α*H*-Gly), 4.56-4.53 (m, 1H, H-3'), 4.12-4.10 (m, 1H, H-4'), 4.07-4.03 (m, 2H, H-5' and H-5''), 4.00-3.94 (m, 2H, OCH_2CH_2S), 3.51-3.46 (m, 1H, α*H*-Ala), 2.83-2.79 (m, 2H, OCH_2CH_2S), 2.31-2.26 (m, 2H, H-2' and H-2''), 1.85 (s, 3H, CH_3-Thy), 1.21 and 1.19 (2 x d, J = 7.0 Hz, 3H, CH_3-Ala); ^{13}C NMR (125 MHz, D_2O) δ = 177.0 (CO-Gly), 173.0 (CO-Ala), 165.5 (C-4), 156.7 (C-2), 135.9 and 135.8 (C-6), 111.3 (C-5), 84.4 (d, $^{3}J_{C,P}$ = 8.4 Hz, C-4'), 84.3 (C-1'), 70.5 and 70.4 (C-3'), 64.5 (d, $^{2}J_{C,P}$ = 5.1 Hz, C-5'), 64.1 (d, $^{2}J_{C,P}$ = 5.5 Hz, OCH_2CH_2S), 55.6 (αC-Gly), 49.2 (αC-Ala), 38.2 (C-2'), 29.6 (d, $^{3}J_{C,P}$ = 7.8 Hz, OCH_2CH_2S), 19.1 (CH_3-Ala), 12.0 and 11.9 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = -0.3; HRMS for $C_{17}H_{27}N_4O_{11}PS$ $[M-H]^-$ calcd.: 525.1062, found: 525.1058.

_N_-Boc-L-Alanyl-D,L-2-_O_-benzylglycine methyl ester (3.31).

To a stirred solution of **3.19** (1.59 g, 5.0 mmol) in dry THF (100 mL), a solution of DABCO (1.35 g, 12.0 mmol) in dry THF (15 mL) was added at -78 °C. After stirring at the same temperature for 10 min, benzyl alcohol (0.62 mL, 6 mmol) was added and the solution was stirred for additional 6 h at -78 °C. The reaction mixture was slowly warmed to room temperature and the stirring was continued for 24 h. The reaction was cooled to 0 °C, quenched with 10% aq. citric acid (100 mL) and all the volatiles were removed *in vacuo*. The residue was extracted with ethyl acetate (3 x 150 mL) and the combined organic layers were washed with 10% aq. citric acid, saturated aq. $NaHCO_3$, water, brine, then dried over Na_2SO_4, filtered and concentrated *in vacuo*. The resulting crude product was purified by column chromatography on silica gel (gradient Acetone/Petroleum ether, 1:99, v/v; 1:19, v/v; 1:9, v/v) to give a diastereomeric mixture of **3.31** (0.82 g, 45%) as a colorless oil. ^{1}H NMR (300 MHz, Acetone-d_6) δ = 8.01 (d, J = 8.6 Hz, N*H*-Gly), 7.39-7.28 (m, Ar-H OBn), 6.26 (br s, 1H, N*H*-Ala), 5.69 and 5.67 (2 x d, J = 9.3 Hz, 1H, α*H*-Gly), 4.71-4.56 (m, 2H, CH_2-OBn), 4.24-4.19 (m, 1H, α*H*-Ala), 3.74 and 3.73 (2 x s, 3H, OCH_3-Gly), 1.42 and 1.41 (2 x s, 9H, ^{t}Bu), 1.36 and 1.35 (2 x d, J = 7.2 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, Acetone-d_6) δ = 174.5 (CO-Ala), 168.9 and 168.8 (CO-Gly), 156.4 (OCONH), 138.7 and 138.6 (1C of OBn), 129.1 (Ar-C), 129.0 (Ar-C), 128.9 (Ar-C), 128.8 (Ar-C), 128.5 (Ar-C), 79.5 (1C-^{t}Bu), 77.7 and 77.6 (αC-Gly), 70.7 and 70.6 (CH_2-OBn), 52.8 (OCH_3-Gly), 51.2 (αC-Ala), 28.5 (^{t}Bu), 18.2 and 18.0 (CH_3-Ala); HRMS for $C_{18}H_{26}N_2O_6$ $[M+H]^+$ calcd.: 367.1863, found: 367.1860.

_N_-Boc-L-Alanyl-L-2-_O_-benzylglycine (3.32).

To a stirred solution of the diastereomeric mixture **3.31** (400 mg, 1.092 mmol) in a 1:1 DMF/H_2O mixture (16 mL) was added at 55 °C subtilisin Carlsberg (20 mg). 1M NaOH (0.61 mL) was then added dropwise by maintaining the pH of the reaction at 7.2 and monitoring by a pH-meter. Upon completion, the volatiles were removed *in vacuo*. The residue was dissolved in 5% $NaHCO_3$ and extracted with ethyl acetate (3 x 50 mL) and the combined organic layers were washed with 5% $NaHCO_3$, 10% aq. citric acid, water, brine, then dried over Na_2SO_4, filtered and concentrated *in vacuo* to give the diastereomeric mixture (320 mg, recovered). The aqueous layer was acidified with 10% citric acid and extracted with ethyl acetate (3 x 50 mL). The combined organic layers were washed with 10% aq. citric acid, water, brine, then dried over Na_2SO_4, filtered and concentrated *in vacuo* to get the pure diastereomer (L,L) **3.32** (54 mg, 14%) as a white solid. ^{1}H NMR (300 MHz, CD_3CN) δ = 7.56 (d, J = 8.6 Hz, N*H*-Gly), 7.36-7.30 (m, Ar-H OBn), 5.67 (br s, 1H, N*H*-Ala), 5.55-5.50 (m, 1H, α*H*-Gly), 4.67-4.53 (m, 2H, CH_2-OBn), 4.10-4.02 (m, 1H, α*H*-Ala), 1.41 (s, 9H, ^{t}Bu), 1.28 (d, J = 7.2 Hz, 3H, CH_3-Ala); ^{13}C NMR (75 MHz, CD_3CN) δ = 175.1 (CO-Gly), 169.2 (CO-Ala), 156.6 (OCONH), 138.4 (1C of OBn), 129.3 (Ar-C), 129.1 (Ar-C), 128.8 (Ar-C), 80.2 (1C-^{t}Bu), 77.6 (αC-Gly), 70.8 (CH_2-OBn), 51.4 (αC-Ala), 28.5 (^{t}Bu), 17.9 (CH_3-Ala); HRMS for $C_{17}H_{24}N_2O_6$ $[M+H]^+$ calcd.: 351.1561, found: 351.1563.

2-*N*,*N*-Diisopropylamino-1,3,2-oxathiaphospholane (3.34).
To a stirred solution of *N,N*-diisopropyl phosphoramidous dichloride (293 µL, 1.588 mmol) and 2-mercapto ethanol (112 µL, 1.588 mmol) in dry diethyl ether (8 mL) was slowly added DIPEA (664 µL, 3.811 mmol) at -20 °C. The reaction mixture was then stirred at the same temperature for 30 min and slowly warmed to room temperature over 2 h. The so formed white precipitate was filtered off under an inert atmosphere and the filtrate was concentrated (bath temp. ~15 °C) and dried *in vacuo* to get crude **3.34** (301 mg, quant.) as a colorless oil. The product was checked for purity using ^{31}P NMR and was immediately used in the next step without any further purification.

3'-*O*-Benzoyl-5'-(2-thio-1,3,2-oxathiaphospholane)-thymidine (3.35).
To a stirred suspension of crude **3.34** (301 mg, 1.588 mmol) and **3.27** (500 mg, 1.444 mmol) in dry CH_2Cl_2 (18 mL) was slowly added a solution of S-ethylthiotetrazole (273 mg, 2.094 mmol) in dry CH_2Cl_2 (5 mL) at 0 °C. The reaction mixture was then stirred at room temperature for 3 h. When the solution became homogeneous, dry elemental sulfur (153 mg) was added and the stirring was continued overnight. The mixture was filtered and the filtrate was concentrated under reduced pressure. The resulting crude residue was purified by column chromatography on silica gel (gradient MeOH/$CHCl_3$, 0:100, v/v; 0.5:95.5, v/v; 1:99, v/v) to give **3.35** (650 mg, 93%) as a pale yellow foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.09 (br s, 1H, N*H*-Thy), 8.07 (d, *J* = 1.0 Hz, 1H, H-6), 8.05-7.46 (m, 5H, OBz), 6.53-6.46 (m, 1H, H-1'), 4.58-4.54 (m, 1H, H-3'), 4.64-4.44 (m, 4H, H-5', H-5'' and OC*H*$_2$CH$_2$S), 4.42-4.39 (m, 1H, H-4'), 3.61-3.52 (m, 2H, OCH$_2$C*H*$_2$S), 2.67-2.60 (m, 1H, H-2'), 2.36-2.23 (m, 1H, H-2''), 2.01 (d, *J* = 1.0 Hz, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 166.3 (CO-OBz), 164.0 (C-4), 150.6 (C-2), 135.4 and 135.3 (C-6), 133.9 (1C-OBz), 129.9 (Ar-C), 128.7 (Ar-C), 112.1 and 112.0 (C-5), 85.1 and 85.0 (C-1'), 83.2 (d, $^3J_{C,P}$ = 8.7 Hz, C-4'), 75.5 and 75.4 (C-3'), 70.9 (d, $^3J_{C,P}$ = 11.3 Hz, OCH$_2$CH$_2$S), 68.1 (d, $^2J_{C,P}$ = 6.6 Hz, C-5'), 37.7 and 37.6 (C-2'), 37.1 (d, $^3J_{C,P}$ = 16.3 Hz, OCH$_2$CH$_2$S), 12.7 and 12.6 (CH$_3$-Thy); ^{31}P NMR (121 MHz, $CDCl_3$) δ = 104.3 and 104.1; HRMS for $C_{19}H_{21}N_2O_6PS_2$ $[M+Na]^+$ calcd.: 507.0420, found: 507.0421.

Thymidine-5'-*O*-phosphorothioate triethylammonium salt (3.36).
To a stirred solution of **3.35** (650 mg, 1.342 mmol) and 3-hydroxypropionitrile (0.46 mL, 6.708 mmol) in dry CH_2Cl_2 (18 mL) was added dropwise DBU (0.22 mL, 1.476 mmol). The reaction mixture was stirred at room temperature for 8 h. It was then concentrated under reduced pressure and the crude residue was purified by flash chromatography on silica gel (gradient MeOH/$CHCl_3$, 1:49, v/v; 1:9, v/v; 1:4, v/v). The fractions containing the phosphorothioate monoester [^{31}P NMR (121 MHz, CD_3OD) δ = 56.6, HRMS for $C_{20}H_{22}N_3O_8PS$ $[M-H]^-$ calcd.: 494.0792, found: 494.0789] were combined, concentrated *in vacuo* and the resulting residue was treated with 25% NH_3 in a sealed tube at 55 °C for 16 h. Ammonia was removed *in vacuo*, the residue was diluted with water and lyophilized. The crude product was purified by preparative RP-HPLC (50 mmol TEAB in 98% H_2O + 2% CH_3CN and 50 mmol TEAB in 50% CH_3CN + 50% H_2O). The collected eluates were freeze-dried repeatedly until constant mass to afford compound **3.36** (497 mg, 72% over two steps) as a white solid. ^{1}H NMR (300 MHz, D_2O) δ = 7.80 (d, *J* = 1.0 Hz, 1H, H-6), 6.34 (app t, *J* = 7.0 Hz, 1H, H-1'), 4.60-4.56 (m, 1H, H-3'), 4.20-4.18 (m, 1H, H-4'), 4.12-4.08 (m, 2H, H-5' and H-5''), 2.36-2.32 (m, 2H, H-2' and H-2''), 1.92 (d, *J* = 1.0 Hz, 3H, C*H*$_3$); ^{13}C NMR (75 MHz, D_2O) δ = 166.2 (C-4), 151.4 (C-2), 137.1 (C-6), 111.4 (C-5), 85.2 (d, $^3J_{C,P}$ = 9.4 Hz, C-4'), 84.7 (C-1'), 71.0 (C-3'), 64.5 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 38.4 (C-2'), 11.3 (CH$_3$-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 50.2; HRMS for $C_{10}H_{15}N_2O_7PS$ $[M-H]^-$ calcd.: 337.0265, found: 337.0262.

Thymidine-5'-O-(*N*-Boc-L-alanyl-D,L-2-aminoglycine methyl ester)-phosphorothioate (3.37). To a stirred solution of compound **3.19** (344 mg, 0.809 mmol) and compound **3.36** (309 mg, 0.970 mmol) in dry DMF (6 mL) was added Et_3N (0.45 mL, 3.234 mmol). The reaction mixture was heated at 37 °C for 4 h. After cooling, DMF was removed under reduced pressure and the resulting residue was dissolved in water and lyophilized. The crude product was purified by preparative RP-HPLC (98% H_2O + 2% CH_3CN and 98% CH_3CN + 2% H_2O). The collected eluates were freeze-dried repeatedly until constant mass to afford a diastereomeric mixture of compound **3.37** (329 mg, 68%) as a white solid. ^{1}H NMR (500 MHz, CD_3OD) δ = 7.74 (s, 1H, H-6), 6.35 (q, J = 6.9 Hz, 1H, H-1'), 5.76-5.69 (m, 1H, α*H*-Gly), 4.50-4.46 (m, 1H, H-3'), 4.16-4.08 (m, 4H, H-4', α*H*-Ala, H-5' and H-5''), 3.79 and 3.78 (2 x s, 3H, OCH_3-Gly), 2.29-2.25 (m, 2H, H-2' and H-2''), 1.96 and 1.94 (2 x s, 3H, CH_3-Thy), 1.47 and 1.46 (2 x s, 9H, ^{t}Bu), 1.35 and 1.34 (2d, J = 7.1 Hz, 3H, CH_3-Ala; merged with Et_3N); ^{13}C NMR (125 MHz, CD_3OD) δ = 175.4 and 175.3 (CO-Ala), 170.8 and 170.7 (CO-Gly), 166.4 (C-4), 157.6 (OCONH), 152.4 (C-2), 138.0 (C-6), 112.0 and 111.9 (C-5), 87.0 (d, $^3J_{C,P}$ = 7.6 Hz, C-4'), 86.2 (C-1'), 80.8 and 80.8 (1C-^{t}Bu), 72.7 and 72.6 (C-3'), 67.0 (C-5'), 55.0 and 54.9 (αC-Gly), 53.6 (OCH_3-Gly), 51.7 and 51.6 (αC-Ala), 40.8 and 40.7 (C-2'), 28.7 (^{t}Bu), 18.2 and 18.0 (CH_3-Ala), 12.7 and 12.6 (CH_3-Thy); ^{31}P NMR (202 MHz, CD_3OD) δ = 13.1; HRMS for $C_{21}H_{33}N_4O_{12}PS$ $[M-H]^-$ calcd.: 595.1480, found: 595.1486.

Thymidine-5'-*O*-(L-alanyl-(-)-2-aminoglycine methyl ester)-phosphorothioate (3.38a) and thymidine-5'-*O*-(L-alanyl-(+)-2-aminoglycine methyl ester)-phosphorothioate (3.38b).
A similar synthetic protocol as the one used for the synthesis of **3.23** was employed for the synthesis of **3.38a-b**, starting from compound **3.37** (160 mg, 0.2293 mmol), thioanisole (33 µL, 0.356 mmol) and TFA (1.5 mL) in CH_2Cl_2 (4.5 mL). Semi-preparative RP-HPLC (98% H_2O + 2% CH_3CN and 98% CH_3CN + 2% H_2O as eluent) was employed for further separation to obtain the TFA salts of pure diastereomers as a ~1:1 ratio (isomer **3.38a** 59.6 mg, isomer **3.38b** 60 mg), as white powders (overall yield **3.38a+3.38b** 119.6 mg, 84%). The isolated products were freeze-dried repeatedly until constant mass. Data for **3.38a**: ^{1}H NMR (300 MHz, D_2O) δ = 7.71 (d, J = 1.0 Hz, 1H, H-6), 6.35 (app t, J = 7.0 Hz, 1H, H-1'), 5.60 (d, J = 15.7 Hz, 1H, α*H*-Gly) 4.61-4.56 (m, 1H, H-3'), 4.20-4.18 (m, 1H, H-4'), 4.16-4.09 (m, 3H, α*H*-Ala, H-5' and H-5''), 3.81 (s, 3H, OCH_3-Gly), 2.42-2.37 (m, 2H, H-2' and H-2''), 1.94 (d, J = 1.0 Hz, 3H, CH_3-Thy), 1.55 (d, J = 7.1 Hz, 3H, CH_3-Ala); ^{13}C NMR (125 MHz, D_2O) δ = 170.1 (d, $^3J_{C,P}$ = 6.2 Hz, CO-Gly), 169.7 (CO-Ala), 166.3 (C-4), 151.5 (C-2), 137.2 (C-6), 111.5 (C-5), 84.9 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 84.8 (C-1'), 70.8 (C-3'), 65.4 (d, $^2J_{C,P}$ = 5.6 Hz, C-5'), 53.6 (αC-Gly and OCH_3-Gly), 48.7 (αC-Ala), 38.4 (C-2'), 15.9 (CH_3-Ala), 11.5 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 15.1; $[\alpha]^{20}_{589}$ = -0.518° (c = 1, CH_3OH); HRMS for $C_{16}H_{25}N_4O_{10}PS$ $[M-H]^-$ calcd.: 495.0956, found: 495.0956. Data for **3.38b**: ^{1}H NMR (500 MHz, D_2O) δ = 7.69 (s, 1H, H-6), 6.36 (app t, J = 7.0 Hz, 1H, H-1'), 5.64 (d, J = 14.5 Hz, 1H, α*H*-Gly) 4.57-4.54 (m, 1H, H-3'), 4.19-4.16 (m, 1H, H-4'), 4.15-4.04 (m, 3H, α*H*-Ala, H-5' and H-5''), 3.75 (s, 3H, OCH_3-Gly), 2.42-2.35 (m, 2H, H-2' and H-2''), 1.88 (s, 3H, CH_3-Thy), 1.53 (d, J = 7.1 Hz, 3H, CH_3-Ala); ^{13}C NMR (125 MHz, D_2O) δ = 169.8 (d, $^3J_{C,P}$ = 6.8 Hz, CO-Gly), 169.7 (CO-Ala), 166.2 (C-4), 151.5 (C-2), 137.3 (C-6), 111.4 (C-5), 84.9 (d, $^3J_{C,P}$ = 8.9 Hz, C-4'), 84.8 (C-1'), 70.6 (C-3'), 65.4 (d, $^2J_{C,P}$ = 5.3 Hz, C-5'), 53.6 (OCH_3-Gly), 53.3 (d, $^2J_{C,P}$ = 3.0 Hz, αC-Gly), 48.8 (αC-Ala), 38.1 (C-2'), 16.0 (CH_3-Ala), 11.5 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 15.1; $[\alpha]^{20}_{589}$ = +0.730° (c = 1, CH_3OH); HRMS for $C_{16}H_{25}N_4O_{10}PS$ $[M-H]^-$ calcd.: 495.0956, found: 495.0959.

General Protocol for the DNA polymerase reaction.

The primer P_1 was purchased from IDT, whilst all the templates T_{1-5} were acquired from Eurogentec. Primer oligonucleotides were labeled with 5'-[γ-33P]-ATP (Perkin Elmer) using T4 polynucleotide

kinase (New England Biolabs) according to the manufacturer's procedure. The labeled primers were additionally purified using Illustra MicroSpin G-25 Column (GE Healthcare) and then annealed with the corresponding template oligonucleotides in a 1:2 molar ratio, by heating the mixture at 75 °C for 5 min, with subsequent slow cooling to room temperature. The DNA polymerisation mixtures contained 125 nM primer-template complex, 1X reaction buffer (supplied with the enzyme), different concentrations of building blocks (125 μM, 500 μM and 1 mM) and 0.01 $U.\mu l^{-1}$ Therminator, 0.01 $U.\mu l^{-1}$ Vent (-exo) or 0.05 $U.\mu l^{-1}$ Klenow (-exo) polymerases (New England Biolabs). The reaction was performed either at 37 °C (mesophilic polymerase) or at 75 °C (thermophilic polymerases) and aliquots were acquired after 15, 30 and 60 min. In the control reaction, 50 μM of the corresponding natural triphosphate TTP was used. All the polymerase reactions were quenched by addition of a double volume of gel loading buffer (90% formamide, 50mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol). Samples were heated either at 75 °C (mesophilic polymerase) or at 90 °C (thermophilic polymerases) for 5 min prior to separation on a 0.4mm 20% denaturing polyacrylamide gel. The bands were then visualized utilizing phosphorimaging (Perkin Elmer).

Steady-state kinetics of single nucleotide incorporation.

Primer 5' /5Cy5/CAGGAAACAGCTATGAC 3' (IDT) was annealed with template 3' GTCCTTTGTCGATACTGATTTT 5' (IDT) in a 1:2 molar ratio by heating the mixture at 75 °C for 5 min, with subsequent slow cooling to room temperature. A series of reactions with different enzyme concentrations and different time points were carried out to achieve the optimum conditions satisfying the 'single completed hit' principle.[20] The final DNA polymerisation mixtures each contained 125 nM primer-template complex, 1X Ex Taq™ reaction buffer (Takara), 2.5 mM $MgSO_4$, building blocks (for modified building blocks the concentration ranged between 15 μM and 2 mM, whilst for natural nucleoside triphosphates the concentration ranged between 10 nM and 2 μM) and 0.005 $U.\mu l^{-1}$ Klenow fragment (New England Biolabs). The reactions were executed at 37 °C and aliquots were acquired after 30 seconds. All polymerase reactions were quenched by adding a double volume of gel loading buffer (90 % formamide, 50 mM EDTA and 0.05% bromophenol blue) and heated at 90 °C for 5 min. The samples were allowed to run on a 1 mm 15% denaturing polyacrylamide gel and gel bands were visualized using Ettan DIGE Imager (GE Healthcare). The gel bands were then measured using ImageQuant TL 1D version 7.0 (GE Healthcare) and the kinetics parameters (V_{max} and K_m) were determined by fitting the data to a non-linear Michaelis–Menten regression utilizing GraphPad Prism Software version 5.0.

Molecular modeling study.

Although a structure of the Klenow fragment of *E. coli* DNA Polymerase I exists in the pdb (2KZM), no ternary complex (enzyme, DNA, triphosphate substrate) was found. Therefore a homology model of the Klenow fragment was created based on the crystal structure of a ternary complex of DNA Polymerase I from Thermos aquaticus (3KTQ). For this model building the I-tasser server was employed. In a second step, DNA, triphosphate and Mg^{2+} ions (copied from 3KTQ) were added to the homology model via a DALI superposition. The triphosphate cytosine base was changed into a thymine (residue TTP) and the complementary base in the template strand was replaced by an adenine. Steric hindrance was removed by vacuum energy minimization using the Amber molecular mechanics suite, while partly restraining the movement of the triphosphate ligand, Mg^{2+} ions and DNA double strand. 3D models of molecules **3.2a** and **3.2b** were superimposed onto the triphosphate using an in house developed flexible superposition program. Both models were then energy minimized in the same way as the structure with the triphosphate substrate. Finally all structures were analysed in the chimera modelling software.[21]

Table 3-2. Overview of interactions between **TTP, 3.2a, 3.2b** molecules and the polymerase.

TTP	**3.2a**	**3.2b**	
-	-16435 kcal/mol	-16287 kcal/mol	Final potential energy after minimization
R432.NH1 / O1A first phosphate	R432.NH1 / O1A oxygen on 1st phosphate	R432.NH1 / O1A oxygen on 1st phosphate	ionic
MG902 /O2A oxygen on 1st phosphate	MG902 / O1A oxygen on 1st phosphate	-	ionic
	MG902 / O2A oxygen on 1st phosphate	MG902 / O2A oxygen on 1st phosphate	ionic
MG901 / O2A oxygen on 1st phosphate	MG901 / O2A oxygen on 1st phosphate	MG901 / O2A oxygen on 1st phosphate	ionic
K435.NZ / O1B oxygen on 2nd phosphate	-	K435.NZ / carboxyl group	ionic
MG901 / O2B oxygen on 2nd phosphate	-	-	Ionic
-	MG901 /carboxyl group	-	Ionic
-	I386.N / C=O of peptide bond	-	Hydrogen bond
-	-	MG901 / C=O of peptide bond	Ionic hydrogen bond
MG901 / O1G oxygen on 3rd phosphate	-	-	ionic
-	F439 / terminal -NH3+	-	cation - pi
Q526.NE2 / O2 on base	Q526.NE2 / O2 of base	Q526.NE2 + Y443.OH / O2 of base	Hydrogen bonds
R345.NH2 / O4' in sugar ring	R345.NH2 / O4' in sugar ring	R345.NH2 / O4'	Hydrogen bonds
Watson-Crick base pair	Watson-Crick base pair	Watson-Crick base pair	hydrogen bonds

3.7. REFERENCES

[1] S. De, E. Groaz, P. Herdewijn, *Eur. J. Org. Chem.* **2014**, *2014*, 2322-2348.

[2] W. D. Kingsbury, J. C. Boehm, R. J. Mehta, S. F. Grappel, C. Gilvarg, *J. Med. Chem.* **1984**, *27*, 1447-1451.

[3] W. D. Kingsbury, J. C. Boehm, D. Perry, C. Gilvarg, *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 4573-4576; S. Y. Hwang, D. A. Berges, J. J. Taggart, C. Gilvarg, *J. Med. Chem.* **1989**, *32*, 694-698; N. J. Hong, Y. T. Park, *Bull. Korean Chem. Soc.* **1993**, *14*, 674-678.

[4] H. Horikawa, T. Iwasaki, K. Matsumoto, M. Miyoshi, *Tetrahedron Lett.* **1976**, *17*, 191-194.

[5] M. Anissimova, L. Yaouancq, M. A. Badet-Denisot, B. Badet, *J. Pept. Res.* **2003**, *62*, 88-95.

[6] C. V. Miduturu, S. K. Silverman, *J. Org. Chem.* **2006**, *71*, 5774-5777.

[7] C. V. Miduturu, S. K. Silverman, *J. Org. Chem.* **2006**, *71*, 5774-5777.

[8] M. Bogenstatter, W. Steglich, *Tetrahedron* **1997**, *53*, 7267-7274.

[9] F. Felluga, W. Baratta, L. Fanfoni, G. Pitacco, P. Rigo, F. Benedetti, *J. Org. Chem.* **2009**, *74*, 3547-3550.

[10] S. De, E. Groaz, L. Margamuljana, M. Abramov, P. Marliere, P. Herdewijn, *Org. Biomol. Chem.* **2015**, *13*, 3950-3962.

[11] C. Ausin, A. Grajkowski, J. Cieslak, S. L. Beaucage, *Org. Lett.* **2005**, *7*, 4201-4204.

[12] S. Iyer, A. C. Hengge, *J. Org. Chem.* **2008**, *73*, 4819-4829.

[13] S. De, E. Groaz, M. Maiti, V. Pezo, P. Marliere, P. Herdewijn, *Tetrahedron* **2014**, *70*, 8843-8851.

[14] R. L. Alexander, S. L. Morris-Natschke, K. S. Ishaq, R. A. Fleming, G. L. Kucera, *J. Med. Chem.* **2003**, *46*, 4205-4208; G. L. Kucera, C. L. Goff, N. Iyer, S. Morris-Natschke, K. S. Ishaq, S. D. Wyrick, R. A. Fleming, L. S. Kucera, *Antiviral Res.* **2001**, *50*, 129-137; N. Schlienger, T. Beltran, C. Perigaud, I. Lefebvre, A. Pompon, A. M. Aubertin, G. Gosselin, J. L. Imbach, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3003-3006.

[15] Y. Ahmadibeni, K. Parang, *Org. Lett.* **2005**, *7*, 1955-1958.

[16] H. H. Zepik, P. Walde, T. Ishikawa, *Angew. Chem. Int. Ed.* **2008**, *47*, 1323-1325.

[17] M. Olesiak, D. Krajewska, E. Wasilewska, D. Korczynski, J. Baraniak, A. Okruszek, W. J. Stec, *Synlett* **2002**, 967-971.

[18] M. Maiti, S. Michielssens, N. Dyubankova, M. Maiti, E. Lescrinier, A. Ceulemans, P. Herdewijn, *Chem. Eur. J.* **2012**, *18*, 857-868; M. Trmcic, F. L. Chadbourne, P. M. Brear, P. W. Denny, S. L. Cobb, D. R. W. Hodgson, *Org. Biomol. Chem.* **2013**, *11*, 2660-2675.

[19] X.-P. Song, C. Bouillon, E. Lescrinier, P. Herdewijn, *Chem. Biodiver.* **2012**, *9*, 2685-2700.

[20] S. Creighton, L. B. Bloom, M. F. Goodman, *Methods Enzymol.* **1995**, *262*, 232-256.

[21] C. A. Brautigam, S. Sun, J. A. Piccirilli, T. A. Steitz, *Biochemistry* **1999**, *38*, 696-704; Y. Li, S. Korolev, G. Waksman, *Embo J.* **1998**, *17*, 7514-7525; A. Roy, A. Kucukural, Y. Zhang, *Nat. Protoc.* **2010**, *5*, 725-738; L. Holm, J. Park, *Bioinformatics* **2000**, *16*, 566-567; R. Salomon-Ferrer, D. A. Case, R. C. Walker, *WIRES Comput. Mol. Sci.* **2013**, *3*, 198-210; E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.

Chapter 4

Sulfonate derived phosphoramidates as active intermediates in the enzymatic primer-extension of DNA and nucleotide delivery in bacteria.

This article is based on an article published by S. De, E. Groaz, L. Margamuljana, M. Abramov, P. Marlière and P. Herdewijn, Org. Biomol. Chem ***2015****, 13, 3950-3962.*

ABSTRACT

Novel unnatural 5'-phosphoramidate nucleosides, have been designed to be evaluated both for delivery of nucleotides and as substrates of DNA polymerases. The mimics feature metabolites such as taurine and a broad range of aliphatic sulfonates coupled through a P-N bond to the 5′-phosphate position of deoxynucleotides, to allow binding interactions in the enzyme active site. The utility of all of the analogues as pyrophosphate mimics was demonstrated in *in vitro* studies for the chain elongation of DNA, using both thermophilic and mesophylic microbial polymerases.

4.1. INTRODUCTION

In biological systems, phosphate esters and anhydrides are amongst the prevailing molecular fragments being released by metabolic transformations requiring good leaving groups, such as elimination and nucleophilic substitution reactions.[1] For instance, DNA polymerization, the polynucleotide synthesis reaction, is an enzymatically-catalysed process, in which a free 3′-hydroxyl nucleophilic group of a *primer*, base-paired to a DNA template, attacks the α-phosphate group of a nucleoside triphosphate with displacement and ensuing hydrolysis of pyrophosphate.[2] This biochemical mechanism might have emerged during evolution as a result of either functional selection or synthetic accessibility.

Our group has an ongoing program devoted to the diversification of nucleic acid metabolism by means of the systematic variation of both structural and enzymatic elements inherent to the genetic propagation of natural biopolymers.[3] One important goal is the identification of novel components that, although not favored by natural selection, may nevertheless exist in nature and follow biochemical pathways orthogonal to natural ones. This may provide an opportunity to implement and propagate *in vivo* a third type of information carrier molecule (XNA), whilst succeeding in the establishment of both genetic and energetic enclaves.

In the search for non-phosphate synthetic alternatives to pyrophosphate we have already demonstrated the leaving group properties of L-aspartic acid and other unnatural amino acids, in the DNA polymerase-catalyzed synthesis of DNA.[4] This selection of substrates was partly motivated by the structural and electronic properties necessary to consent accommodation and recognition within the enzyme catalytic site, but most importantly by their significance as non-toxic intermediates of metabolic steps in living cells.

In an extension of the substrate scope of the DNA polymerization reaction, we became interested in 5'-phosphoramidate deoxyribonucleoside analogues conserving the anionic residues for key binding interactions with polymerases in the form of aliphatic sulfonate-based functionalities.[5] The design of such unnatural mimics of triphosphates is defined by the abundance of simple organic molecules such as taurine (2-aminoethanesulfonic acid) and L-cysteic acid in natural environments and nutrients and their contribution as sources of mineral sulfur to microbial growth.[6] Organic sulfonates are ionized at physiological pH with very low pK_a values, thus their cellular uptake cannot occur by passive diffusion, but are rather taken up across bacterial membranes by complex transport systems and actively degraded by the action of oxygenolytic enzymes over desulfonation pathways. Typically, this process releases sulfite as degradative intermediate, which then enters the sulphur assimilation cycle. This is one of the essential conditions that a good leaving group candidate must fulfil in order for nucleotide incorporation to be irreversible.

It is well documented in the literature that under conditions of inorganic sulfate and cysteine

starvation, various gene clusters are expressed in *E. coli*, allowing the alternative utilization of aliphatic sulfonates.[7] Two operons, *tau*ABCD[8] and ssuEADCB,[9] have been identified which encode for distinct ABC-type permeases, thus promoting uptake of taurine and other alkanesulfonates respectively, while the TauD gene expresses for the protein alpha-ketoglutarate-dependent dioxygenase responsible for the liberation of sulphite from taurine.[10] Therefore, given our interest in developing an uptake system for nucleotides in bacterial cells, the synthesis of these molecules was also deemed important in view of their potential use as substrates for cellular delivery within a sulfur-based nutritional selection system.

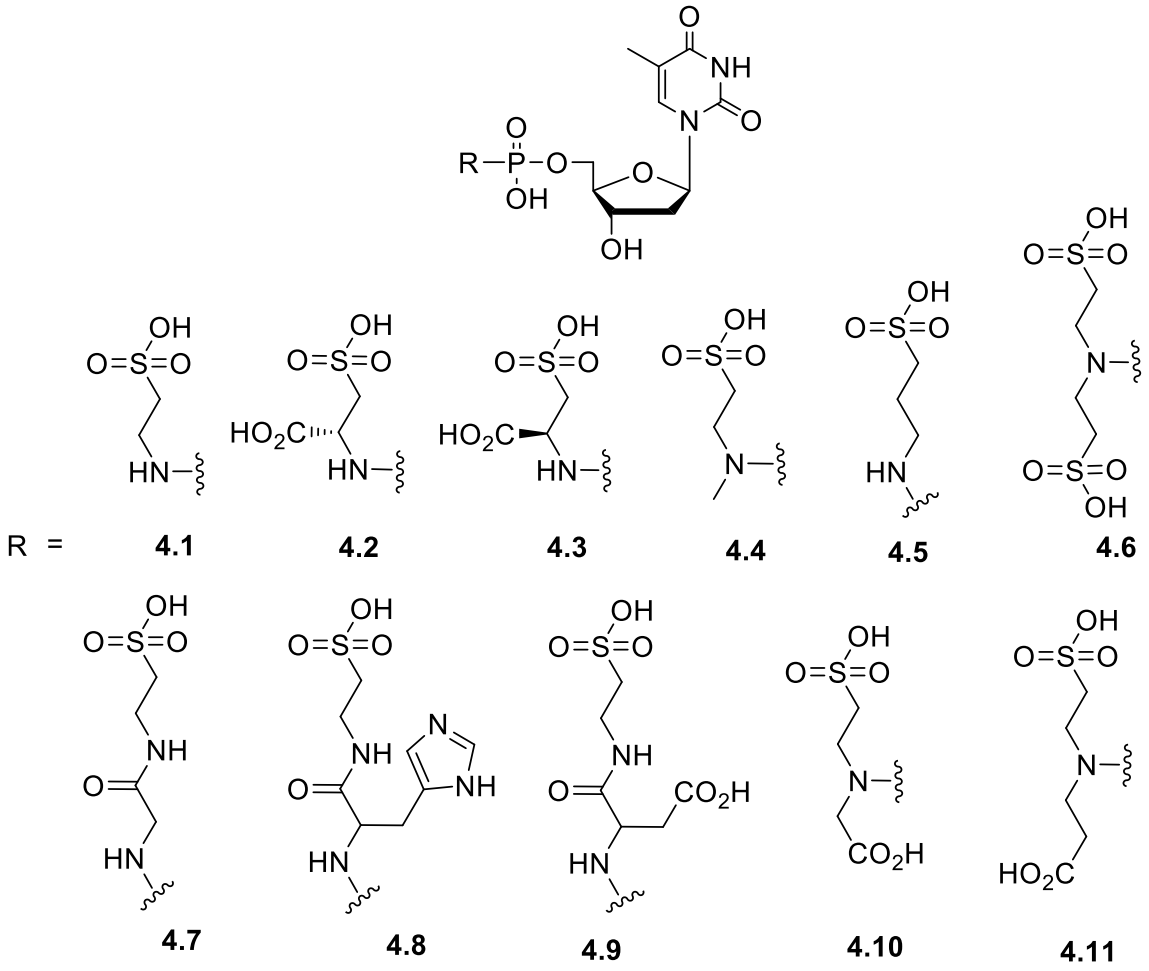


Figure 4-1. Overview of sulfonate-containing phosphoramidate analogues of thymidine.

Herein, we report novel straightforward syntheses of taurine, L-cysteic acid and a wide range of relevant sulfono-modified phosphoramidate nucleotide analogues (examples of TMP derivatives are shown in Figure 4-1), as well as an account of the efficiency with which various mesophilic and thermophilic microbial polymerases accept them as substrates.

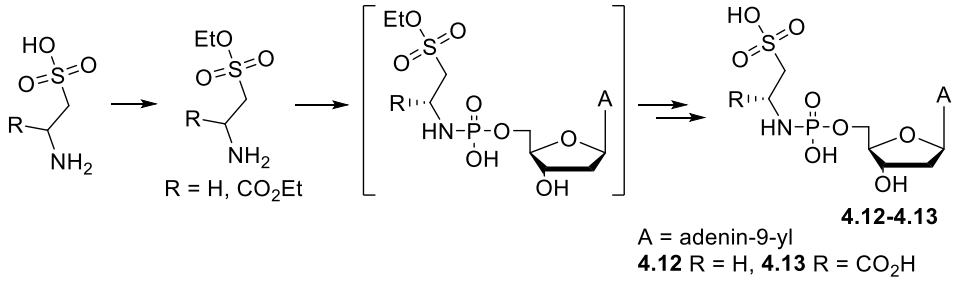
Along with the biochemical relevance, the knowledge of the extent to which molecular diversity can be manipulated but still accepted within the polymerase active site will provide fundamental parameters in the selection and design of new synthetic genetic variants.

4.2. RESULTS AND DISCUSSION

In an early communication, we established a synthetic approach to sulfonate-containing phosphoramidates Tau-dAMP **4.12** and L-Cys-dAMP **4.13**, which centred upon an EDC-mediated coupling reaction between deoxyadenosine monophosphate (dAMP) and the relevant aminoalkylsulfonate motifs.[11] However, protection of both sulfonic and carboxylic acids as ethyl ester groups was necessary to prevent formation of undesired side-products (Scheme 4-1).

In addition, it was anticipated that in order to accomplish the synthesis of higher functionalized substrates, multiple selective protection and deprotection steps of acidic groups would be cumbersome, and most likely produce the novel phosphoramidate products in inadequate yields.

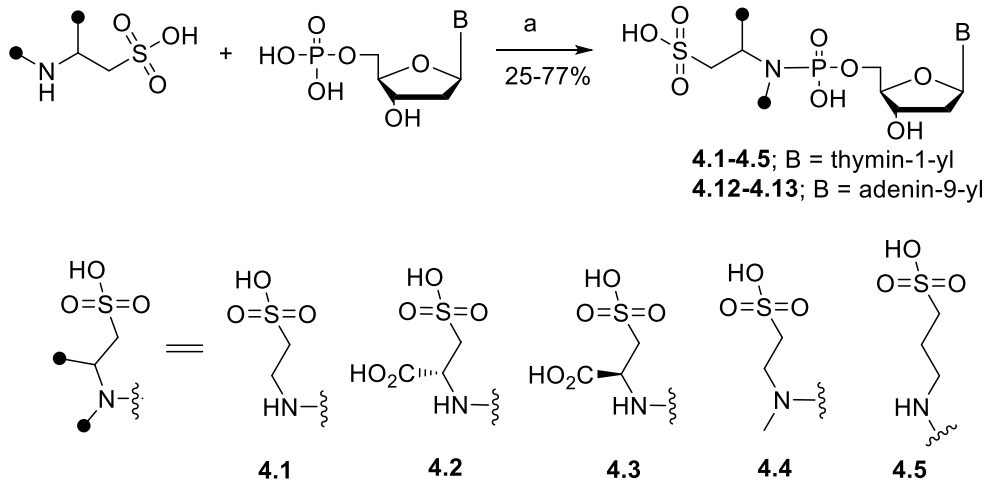
Surprisingly, the replacement in our existing method of EDC by DCC in ^{t}BuOH-H_2O as coupling agent proved to be an alternative satisfactory procedure, even so in the presence of unprotected sulfonic acid functionalities. The resulting adenosine 5'-sulfono-phosphoramidates **4.12** and **4.13** were obtained in a single step with increased yields, provided that an excess of Et_3N (5 eq.) was added to the reaction mixture in order to mask the charges of the sulfonic acid groups. Similarly, sulfono-phosphoramidates **4.1**, **4.4** and **4.5** were isolated in good yields (72-77%) from thymidine monophosphate (TMP) and either taurine, *N*-methyl taurine or homo-taurine as starting materials (Scheme 4-2). This methodology was also applied to the preparation of L-Cys-dTMP **4.2** and D-Cys-dTMP **4.3** in the presence of both free sulfonic and carboxylic acids, albeit with reduced coupling efficiencies (25-30% yields) (Scheme 4-2).



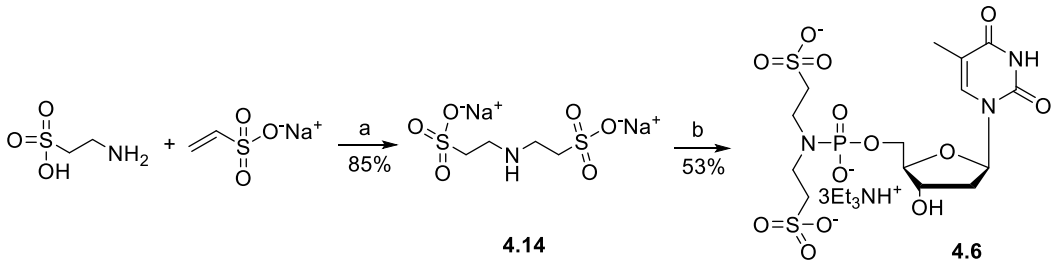
Scheme 4-1. Early route towards sulfono-phosphoramidates *via* ester protection of sulfonic acid.

The same DCC-mediated protocol was also valuable to generate bis-sulfonate phosphoramidate **4.6** after coupling between the triethylammonium salt of TMP and the secondary amine *N*,*N*-di-2-aminoethanesulfonate **4.14**, afforded in turn from the Michael addition of taurine to sodium vinyl sulfonate (Scheme 4-3).

In order to provide further evidence to the geometry accepted by the active site,[12] we decided to prepare dipeptide-like phosphoramidate thymidine analogues bearing various amino acids interconnecting the taurine unit with the amidate bond, as shown in Scheme 4-4. Peptidic bonds can be degraded *in vivo* by hydrolase enzymes releasing amino acid and taurine.



Scheme 4-2. Synthesis of sulfono-phosphoramidate analogues of thimydine **4.1**-**4.5** and adenosine **4.12** and **4.13**. (a) TMP or dAMP (Na or Et_3N salt), DCC, Et_3N, ^{t}BuOH: H_2O (4:1), 90 °C, 4-6 h.

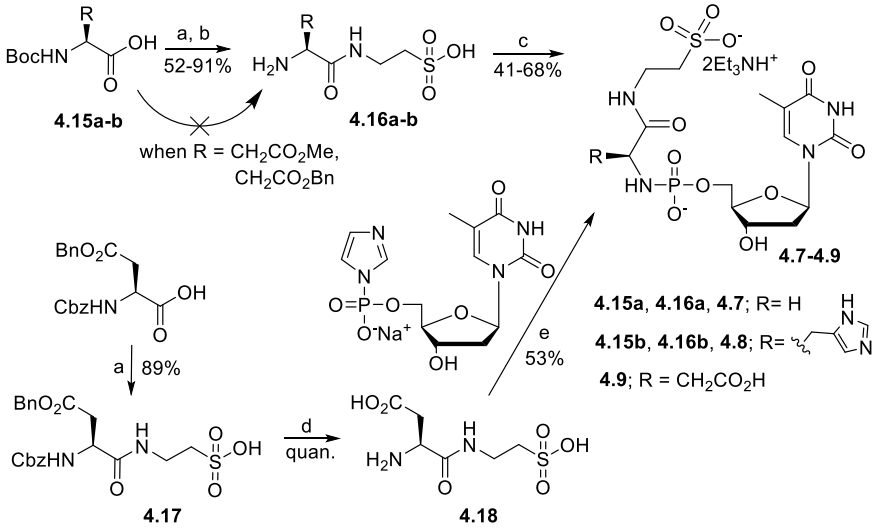


Scheme 4-3. Synthesis of thymidine phosphoramidate analogue **4.6**. (a) $NaHCO_3$, H_2O, reflux, 72 h; (b) TMP, DCC, Et_3N, ^{t}BuOH: H_2O (4:1), 90 °C, 5.5 h.

The first pyrophosphate mimic of this series containing the neutral residue glycine, compound **4.16a**, was accessible through standard coupling conditions in the presence of DCC/NHS in a mixture of dioxane and water, followed by Boc-removal using IR-120B (H^+) resin.[13] When taurine was reacted with a non-neutral amino acid, the same two-steps sequence was found to be less rewarding, producing compound H-L-His-Tau-OH **4.16b** in modest yield. All attempts to induce coupling of taurine to aspartic acid, however, met with failure producing only undesired side products (uncharacterized), with both methyl and benzyl featuring as protecting groups of the β-carboxyl functionality. Thus, compound H-L-Asp-Tau-OH **4.18**, was alternatively obtained in two-steps starting from protected Z-L-Asp (OBn)-OH and taurine, followed by hydrogenolysis. The final modified nucleosides were prepared under standard DCC coupling conditions, as outlined in Scheme 4-4.

As the reaction between TMP and **4.18** led to the isolation of phosphoramidate **4.9** in low yield, which we believe to be due to the presence of a free carboxylic acid group in compound **4.18**, we opted for a different route for this latter substrate. The 5'-phosphate group was thus pre-

activated by conversion into a better phosphoraimidazolide leaving group, which allows for enhanced nucleophilic attack of amines even under mild conditions.

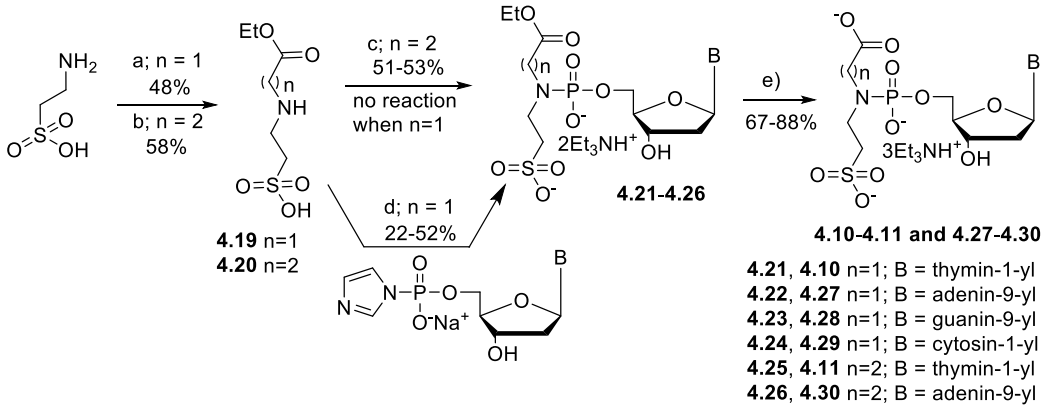


Scheme 4-4. Synthesis of taurine-amino acid phosphoramidate nucleoside analogues **4.7**-**4.9**. (a) (i) NHS, DCC, THF, 0 °C to r.t., 2 h, (ii) Taurine Na-Salt, Dioxane-H_2O (1:1), r.t., 14 h; (b) IR-120B resin (H^+ form); (c) TMP-Na salt, DCC, Et_3N, ^tBuOH:H_2O (4:1), 90 °C, 4.5 h; (d) 10% Pd/C, H_2, EtOH-H_2O (5:1), r.t., 16 h; (e) Et_3N, DMF, 35 °C, 168 h.

N-Acetic acid-2-aminoethane sulfonic acid and *N*-propionic acid-2-aminoethane sulfonic can mimic respectively iminodiacetic acid (IDA) and iminodipropionic acid (IDP), previously reported by our group as very good pyrophosphate mimics for DNA polymerization.[14, 15] Therefore, a series of 2'-deoxynucleoside-5'-*O*-[*N*-(alkyl acid)-2-aminoethane sulfonic acid] phosphoramidates was prepared according to a three-step synthetic methodology (Scheme 4-5), starting from taurine without the need for protection of the sulfonic acid group. The formation of *N*-(ethyl acetate)-2-aminoethane sulfonic acid **4.19**, following the nucleophilic diplacement of the bromo atom in ethylbromoacetate by taurine, occurred only upon heating, whilst *N*-(ethyl acetate)-2-aminoethane sulfonic acid **4.20** resulted from the Michael addition of taurine to ethyl acrylate at ambient temperature over seven days.

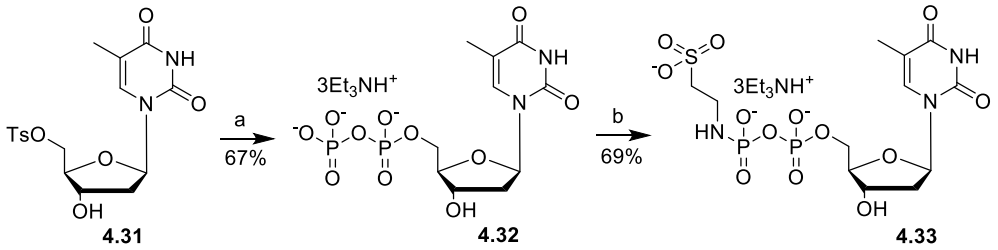
As it might be speculated that DCC-mediated phosphoramidate coupling of secondary amines would prove less effective than that of primary amines, we thought it necessary to employ protected carboxylic ester derivatives in order to improve the reaction outcome in terms of isolated yields. Accordingly, reactions of TMP or dAMP thiethylammonium salts with secondary amine **4.20** were both reasonably efficient to yield respectively compound **4.25** and dAMP-analogue **4.26**. However, to our surprise, the secondary amine *N*-(ethyl acetate)-2-aminoethane sulfonic acid **4.19** failed to react under this condition. Thus, for the synthesis of compounds **4.21**-**4.24** bearing all four nucleobases, we adopted the phosphoraimidazolide- based protocol described above, which delivered the T, A and C analogues **4.21**, **4.22**, and **4.24** in moderate

yields (38-52%) and G analogue **4.23** in modest yield (22%). Ultimately, the deprotection of the ethyl ester group was carried out by treatment with a dilute solution of sodium hydroxide (0.4 M), and subsequent HPLC purification furnished the pure unprotected analogues **4.10**-**4.11** and **4.27**-**4.30** for incorporation studies.



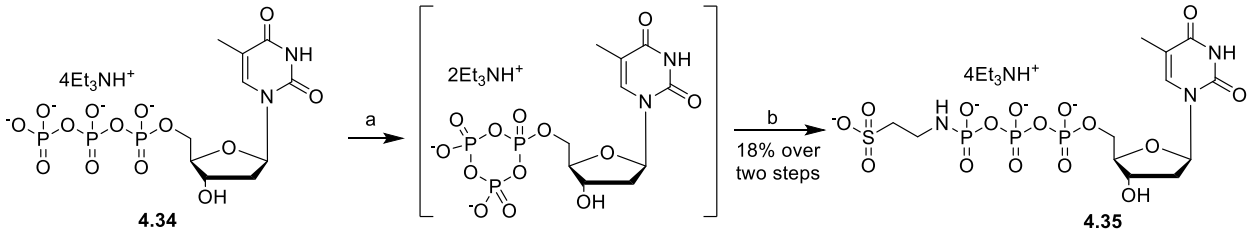
Scheme 4-5. Synthesis of sulfono-phosphoramidate nucleoside analogues **4.10**-**4.11** and **4.27**-**4.30**. (a) Ethylbromo acetate, $NaHCO_3$, Dioxane/H_2O, 70 °C, 24 h; (b) Ethyl acrylate, $NaHCO_3$, EtOH-H_2O (1:1), r.t., 168 h; (c) dTMP-Et_3N salt, DCC, Et_3N, ^{t}BuOH-H_2O (4:1), 90 °C, 6 h; (d) Nucleoside-Imidazolide Na-Salt, Et_3N, DMF, 35 °C, 168 h; (e) 0.4 M NaOH in MeOH:H_2O (4:1 v/v), r.t., 3 h.

Next, Tau-dTDP β-phosphoramidate **4.33** was prepared according to the DCC-mediated phosphoramidate coupling between taurine and the triethylammonium salt of thymidine diphosphate **4.32**. Compound **4.32** was derived in turn from 5'-*O*-tosyl derivative **4.31**,[16] upon treatment with tris(tetra-n-butylammonium) hydrogen pyrophosphate, acting as a nucleophile to displace the tosyl group as reported by Davisson *et. al* (Scheme 4-6).[17]



Scheme 4-6. Synthesis of taurine diphosphoramidate analogue **4.33** of thymidine. (a) Tris (tetra-n-butylammonium) hydrogen pyrophosphate, CH_3CN, r.t., 24 h; (b) Taurine, DCC, Et_3N, ^{t}BuOH:H_2O (4:1), 85 °C, 4.5 h.

Finally, the synthesis of Tau-dTTP γ-phosphoramidate **4.35** was accomplished through initial conversion of dTTP-Et_3N salt **4.34** to thymidine-5'-trimetaphosphates by dehydration with DCC,[18] followed by ring opening of thymidine-5'-trimetaphosphate with an excess of taurine (Scheme 4-7).



Scheme 4-7. Synthesis of taurine triphosphoramidate analogue **4.35** of thymidine. (a) DCC, DMF, r.t., 3.5 h; (b) Taurine, MeOH/DMF, r.t., 6 h.

4.3. CHAIN ELONGATION EXPERIMENTS (performed by Lia Margamuljana)

The individual competency of chemically-modified sulfono-phosphoramidate nucleosides **4.1-4.11**, **4.27-4.30**, **4.32-4.33** and **4.35** to perform as active intermediates in the DNA polymerization reaction was assessed by comparing the activities of different bacterial DNA polymerases, including Therminator, Vent (exo-) and the Klenow fragment (exo-) of *E. coli* DNA Polymerase I. The last two mutants lacking 3′→ 5′proofreading activity were included in the study as preferred for affording high yield primer extension reactions in the presence of poor substrates, by suppressing simultaneous synthesis and editing.

Multiple nucleotides incorporation efficiencies in primer extension were estimated by means of a polyacrylamide gel-based template dependent assay, using a 5'-radiolabelled γ-^{33}P DNA primer annealed in turn to **T1-5** templates (Table 4-1). For each enzyme, reactions were also carried out with natural deoxynucleoside triphosphates (dNTP) as positive controls.

Table 4-1. Overview of primer and templates sequences used in the incorporation experiments.

Primer P1: 5' /5Cy5/CAGGAAACAGCTATGAC 3'
Template T1: 3' GTCCTTTGTCGATACTGAAAAA 5'
Template T2: 3' GTCCTTTGTCGATACTGTTTTTTT 5'
Template T3: 3' GTCCTTTGTCGATACTGGGGGG 5'
Template T4: 3' GTCCTTTGTCGATACTGCCCCC 5'
Template T5: 3' GTCCTTTGTCGATACTGCAAAA 5'

The first set of compounds to be examined as potential triphosphate mimics, were adenine sulfonate-phosphoramidates **4.12-4.13** and thymidine analogues **4.1-4.5**. Preliminary results of single nucleotide incorporation experiments for analogues Tau-dAMP **4.12** and L-Cys-dAMP **4.13** had shown modest incorporation efficiencies into a growing DNA chain, as catalyzed by

HIV-1 reverse transcriptase[11] or family C *E. coli* Pol III.[19] However, it is recognized that phylogenetically distinct classes of enzymes might show different substrate-specificities, despite sharing similar polymerization domains and mechanism. As evidence of a dynamic effect triggered by multiple incorporations on the progression of the reaction emerged in the course of previous studies,[19] all compounds were directly subjected to elongation tests.

While Vent (exo-) was not able to accept adenine analogues Tau-dAMP **4.12** and L-Cys-dAMP **4.13** or thymidine analogues Tau-dTMP **4.1**, L-Cys-dTMP **4.2** and D-Cys-dTMP **4.3**, 6% and 14% formation of the P + 1 strand was observed at 1mM (60 min) with Klenow (exo-) for cysteic acid-containing thymidine phophoramidates **4.2** and **4.3**, and incorporation efficiencies with the same polymerase were relatively higher for adenine analogues **4.12** and **4.13**, showing 5% (P + 2) and 6% (P + 4) respectively. However, the thermostable polymerase Therminator could use all those compounds as substrates, in line with the significant propensity of this replicative enzyme to accept modified nucleotides substrates. We have observed 75% (P + 1) and 6% (P + 2) for **4.1**, 68% (P + 1) and 6% (P + 2) for compound **4.2**, 70% (P + 1) and 16% (P + 2) for compound **4.3**, 18% (P+1) for compound **4.12** and complete conversion to (P + 1) strand, 58% of (P + 2), 28% of (P + 3), 9% of (P + 4) and up to 2% of (P + 5) strand for compound **4.13**.

Within the series of analogues containing thymine as nucleobase, *N*-methyl taurine derivative *N*-Me-Tau-dTMP **4.4** showed to be incorporated into the primer template duplex more efficiently than compounds bearing taurine, D- or L-cysteic acid, with 14% conversion up to (P + 2) strand using Vent (60 min, 1mM) and 17% conversion up to (P + 3) using Therminator (60 min, 1mM), whilst homotaurine-dTMP **4.5** showed only 6% and 20% conversion to (P + 1) strand with Vent and Therminator (60 min, 1mM) respectively. The leaving group properties of bis-taurine in derivative **4.6** was comparable to that of *N*-methyl taurine, showing formation of 15 % of (P + 2) and 2% of (P + 3) strand with Vent and 21 % of (P + 2) and 3% of (P + 3) strand with Therminator at 60 min (1mM).

The acceptance of taurine-amino acid containing derivatives Tau-Gly-dTMP **7**, Tau-His-dTMP **4.8** and Tau-Asp-dTMP **4.9** in primer extension reactions by Vent (exo-) and Therminator polymerases was disappointing, with the exception of aspartate analogue **4.9** which showed 17% P + 1 and 2% P + 2 after 60 min (1mM) with Therminator. Since the individual aminoacids composing those structural units are proved to be good leaving groups, it can be concluded that the increased linear size of those compounds does not allow accommodation in the enzymatic pocket, and the results obtained for aspartate analogue **4.9** might be due to induction from the β-carboxylic functionality.

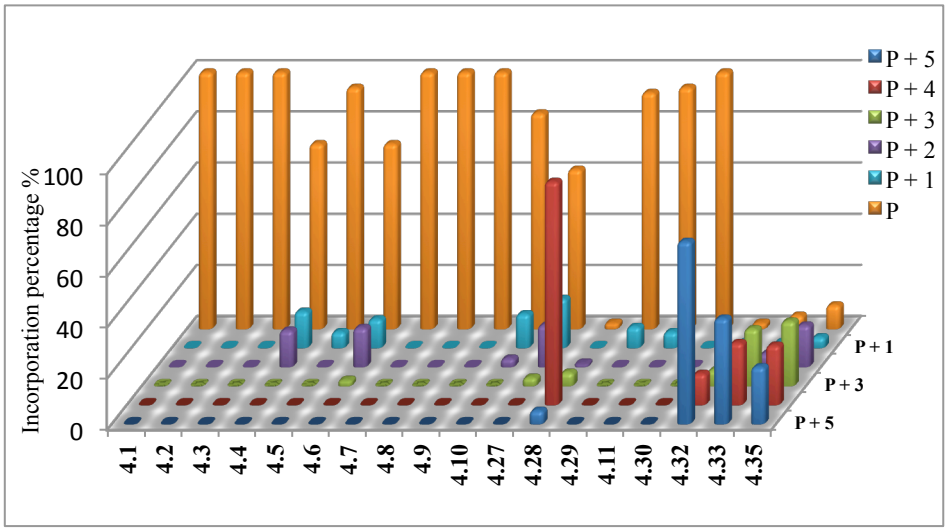


Figure 4-2. Chain elongation efficiency of compounds **4.1-4.11**, **4.27-4.30**, **4.32-4.33** and **4.35** by Vent (exo-) polymerase (0.01 U.µl-1), at 1mM conc. after 60 min.

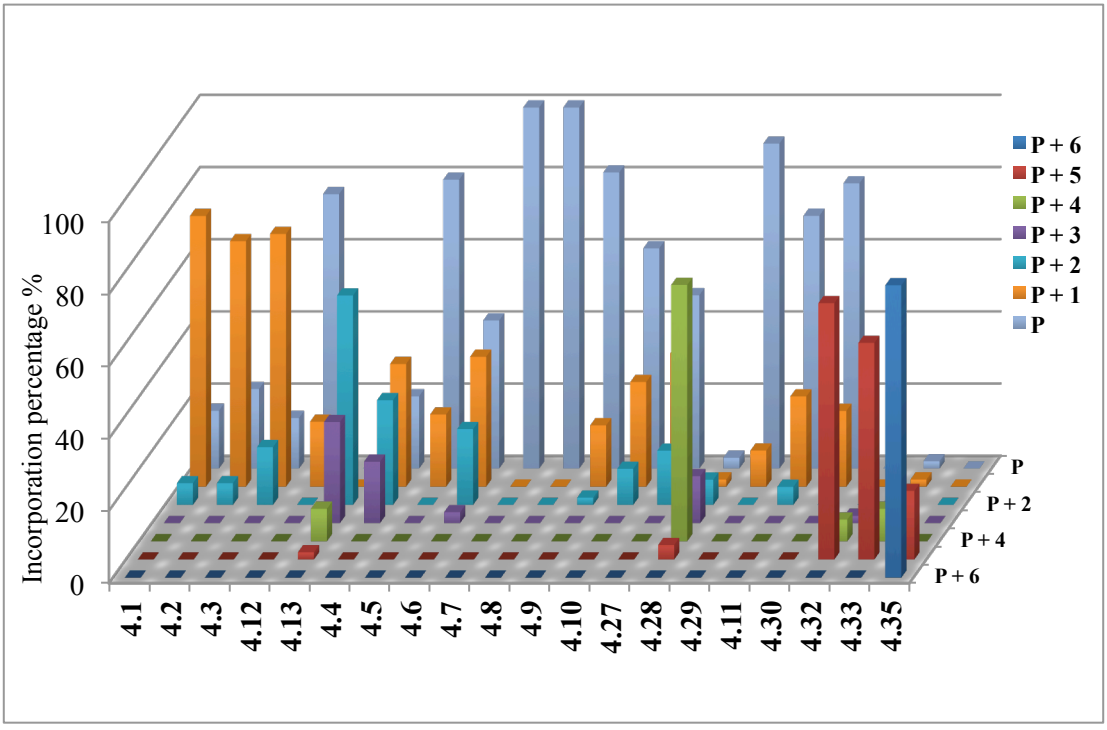


Figure 4-3. Chain elongation efficiency of compounds **4.1-4.13**, **4.27-4.30**, **4.32-4.33** and **4.35** by Therminator polymerase (0.01 U.µl-1), at 1mM conc. after 60 min.

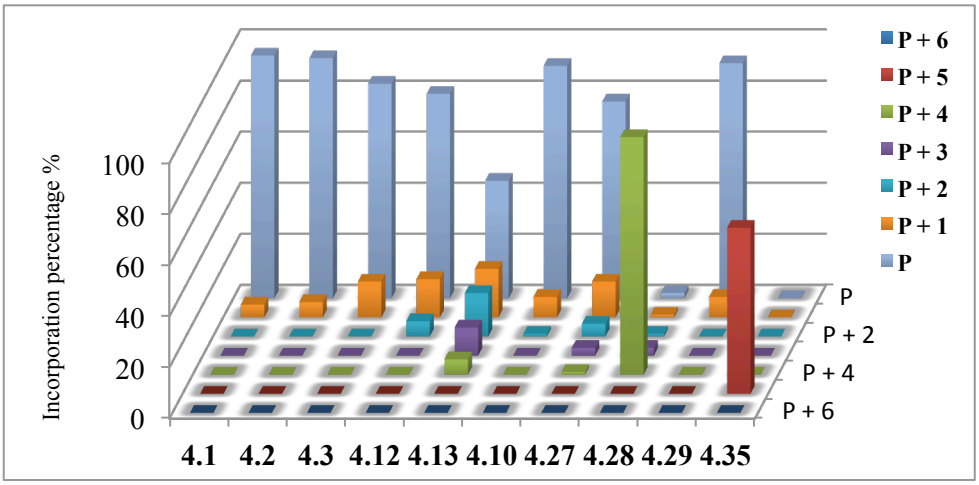


Figure 4-4. Chain elongation efficiency of compounds **4.1-4.3**, **4.10**, **4.12-4.13**, **4.27-4.30** and **4.35** by Klenow fragment polymerase (exo-) (0.05 U.µl-1), at 1mM conc. after 60 min.

The next analogues to be studied were the phosphoramidates derived from taurine-*N*-acetic acid and taurine-*N*-propionic acid. Previously, we observed a preference for incorporation for purines over pyrimidines with various polymerases with the order A ≥ T = G > C, so we initially wished to compare with a comparable efficiency, taurine-*N*-acetic acid adenine nucleoside **4.27** was a better substrate for thermophilic polymerases than taurine-*N*-propionic acid adenine analogue **4.30**. In particular, for the thymidine analogue **4.10**, only 3%, 1% and up to 10% formation of P + 2 strand in 60 min at 1mM was observed with Vent, Klenow and Therminator respectively, whereas the incorporation properties of adenine analogue **4.27** were better than the thymidine analogue, showing up to 3% P + 3 strand with Vent, 15% of P + 2 with Therminator and 1% of P + 4 strand with Klenow. In view of these observations, it was considered important to additionally explore the influence of the other nucleobase moieties on the incorporation efficiency. To this end, we synthesized and investigated deoxynucleoside monophosphate analogues of G and C nucleobases **4.28** and **4.29**, modified at the 5'-position with the selected taurine-*N*-acetic acid as leaving group using Vent (exo-), Therminator and Klenow as polymerases. As anticipated, the cytidine analogue **4.29** showed the lowest efficiency, giving rise to the formation of 8% of P + 1 with Vent and Klenow and 10 % of (P + 1) only with Therminator. Surprisingly, guanidine analogue **4.28** was added into a growing DNA chain more efficiently than any other nucleotide, with a final base ranking order G > A ≥ T > C for this set of derivatives, showing formation of 87% of (P + 4) and up to 5% of full length (P + 5) strand with Vent, up to 93% of (P + 4) with Klenow and 71% of (P + 4) and 4% of (P + 5) strand with Therminator (Figure 5). The rationale behind the base-specificity of DNA polymerases is uncertain. In a distinct literature report, it has been shown that kinetics of incorporation could be affected by the diversification of the 3'-flanking neighbouring base on the template and between

pairing of dCTP or dTTP opposite to O^6-methylguanine.[20]

Ultimately, the elongation ability of diphosphate analogue tau-dTDP **4.33** was excellent, giving up to 41% full-length chain (P + 5) formation at 1 mM in 60 min with Vent and 9 % full length chain (P + 5) formation even at 0.125 mM in 15 min with Therminator. These findings were thought to be due to the presence of the diphosphate group, thus as additional evidence we compared compound **4.33** with natural 5'-thymidine diphosphate **4.32** (dTDP). Enzymatic results pointed to a slightly reduced incorporation efficiency for tau-dTDP in comparison to dTDP. Unsurprisingly, the last compound to be evaluated, Tau-dTTP **4.35**, was found to be as efficient as the natural substrate dTTP using both Therminator and Klenow although less efficient in the presence of Vent (exo-), showing only 22% formation of P + 5 strand at 1mM concentration.

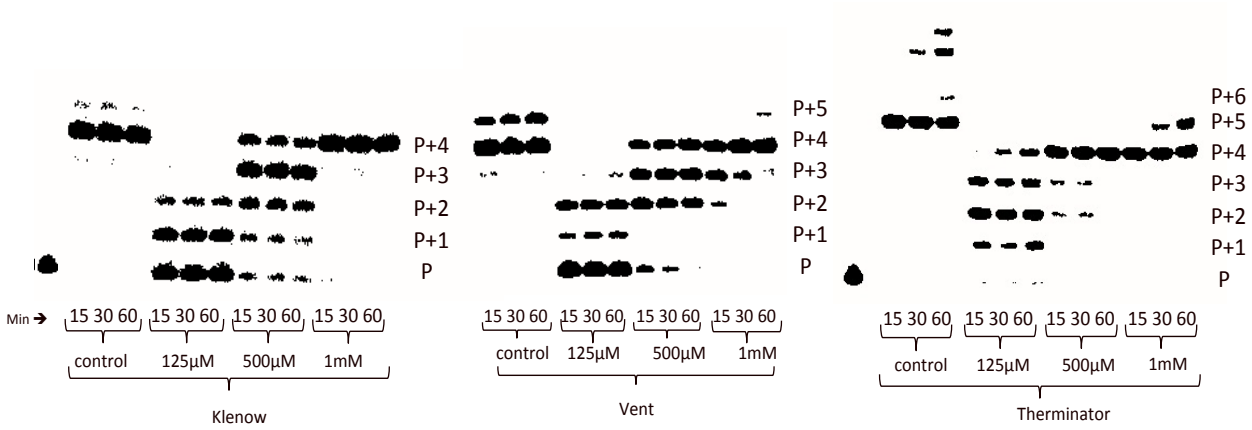


Figure 4-5. Profile of chain elongation of the control **dGTP** and compound **4.28** as substrate into the P1:T4 by three different DNA polymerase; Klenow (0.05 U.µl-1), Vent (0.01 U.µl-1) and Therminator (0.01 U.µl-1).

Table 4-2 Steady–state kinetics of single nucleotide incorporation into P1T5 by Klenow fragment (0.001 U.µl-1)

Substrate	V_{max} [nM.min^{-1}]	K_m [µM]	V_{max} / K_m [x 10^{-3} min^{-1}]
dGTP	58.10 ± 3.23	0.089 ± 0.015	652.80
Compound **4.28**	62.55 ± 3.30	67.34 ± 11.00	0.93

As the best results were seen with modified guanidine substrate **4.28** and Klenow fragment (exo-), incorporation and subsequent extension of compound **4.28** into primed-template P_1:T_5 were quantitatively determined, based on the Single Complete Hit model, in comparison to its natural nucleoside triphosphate (dGTP). The steady-state kinetic analysis, the results of which are

summarized in Table 4-2, indicates a V_{max} value for compound **4.28** 1.07 fold higher than that of natural dGTP, but a K_m value 758 fold higher, with a V_{max}/K_m ratio of compound **4.28** 702-fold lower than the natural substrate. Although the relevant nucleobase (adenine) and DNA enzyme (HIV-1 RT) previously employed in the analysis of bio-isosteres iminodiacetic acid (IDA) and iminodipropionic acid (IDP) were different, we can nevertheless conclude that the leaving group properties of taurine-*N*-acetic acid is comparable to IDA (IDA-dAMP resulted 940-fold lower than the natural substrate) and slightly lower than IDP (IDP-dAMP was 83-fold lower than the natural substrate).

4.4. CONCLUSION

A novel series of modified deoxyribonucleosides bearing sulfonate or both carboxylate and sulfonate functional groups, connected through a phosphoramidate linker to the 5'-position were successfully prepared. Following multiple incorporation screening of various microbial polymerases (as reviewed in Figures **4.2**, **4.3** and **4.4**), the A family Klenow fragment exonuclease-free polymerase mutant was found to be the most effective polymerase, giving the best elongation rates, while B family thermostable Therminator was confirmed as possessing the broader substrate-specificity. The development of new leaving groups for DNA polymerization can be a valuable tool for synthetic biology, for which the mesophylic Klenow polymerase from *E. coli* was tested and should aid the search for novel reagents able to perform DNA synthesis at high temperature for polymerase chain reaction (PCR) for which thermostable DNA polymerases were tested.

4.5. EXPERIMENTAL SECTION

General information

For all reactions, analytical grade solvents were used. All moisture-sensitive reactions were carried out under an argon or a nitrogen atmosphere in oven-dried glassware (135 ºC). Reaction temperatures are reported as bath temperatures. If not mentioned, all amino acids used were L-isomer. Precoated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ = 254 nm). Products were purified by flash chromatography on ICN silica gel 63-200, 60 Å. All final compounds were purified by preparative RP-HPLC (XbridgeTM Prep C18 5µm OBD 19 x 150 mm column) or Ion-exchange resin (Source 15 Q). All the elution methods use MeCN/H_2O gradients for RP-HPLC or TEAB-H_2O gradient for Ion-exchange. ^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on Bruker Avance 300 MHz, 500 MHz or 600 MHz spectrometers. For all final compounds, 1H and 13C resonance assignments were made using 2D NMR correlation experiments (COSY, gHSQC and gHMBC). For sake of clarity, NMR signals of protons and carbons for sugar and base moieties are indicated with and without a prime, respectively. Chemical shifts were referenced to residual solvent signals at δ H/C 7.26/77.00 ($CDCl_3$), 3.31/49.10 (CD_3OD) and 2.50/39.50 (DMSO-d_6) relative to TMS as internal standard. Coupling constants are expressed in hertz (Hz). Splitting patterns are reported as s (singlet), d

(doublet), t (triplet), q (quartet), m (multiplet), br (broad) and app (apparent). High-resolution mass spectra were acquired on a quadruple orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 μL/min and spectra were obtained in positive or negative ionization mode with a resolution of 15000 (FWHM), using leucine enkephalin as lock mass.

General procedure for phosphoramidates synthesis (A)

The relevant 2'-deoxynucleoside-5'-monophosphate (1 eq.) and amino derivative (3-5 eq.) were suspended in a 4:1 ^{t}BuOH/H_2O mixture (~20 mL/mmol). Triethylamine (5-10 eq.) was then added to the suspension to facilitate dissolution, followed by DCC (2-5 eq.). The reaction mixture was heated at 90 ºC for 3-6 h. The reaction progress was monitored by TLC (iPrOH:H_2O:Et_3N/NH_3 10:2:2). Upon completion, the reaction mixture was cooled to room temperature and the solvent was removed by rotary evaporation. The residue was resuspended in water (100 ml), washed with diethyl ether (3 x 100 ml) and the aqueous phase was lyophilized. The resulting crude material was purified by column chromatography on silica gel using the following gradient IPA:H_2O:NH_3 20:1:1, v/v/v; 15:1:1, v/v/v; 10:1:1, v/v/v (Et_3N was used instead of NH_3 for carboxylic ester containing compounds, to avoid amide formation), to provide the desired nucleoside phosphoramidates as salts. Semi-preparative RP-HPLC (50 mmol TEAB in H_2O + 2% ACN and 50 mmol TEAB in 50% H_2O + 50% ACN) was employed for further purification to obtain the pure products (except for compounds **4.1**, **4.2**, **4.3** and **4.8**). The residues were lyophilized and freeze-dried repeatedly until constant mass.

Thymidine-5'-taurine phosphoramidate ammonium salt (4.1)

The ammonium salt of compound **4.1** was obtained as a white solid (0.27 g, 72%) following the general procedure **(A)**, starting from 2'-deoxythymidine-5'-monophosphate (TMP) disodium-salt (0.3 g, 0.82 mmol), taurine (0.31 g, 2.46 mmol), triethylamine (0.57 mL, 4.09 mmol), DCC (0.51g, 2.46 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (15 mL) at 90 ºC for 3 h. ^{1}H NMR (300 MHz, D_2O) δ = 7.69 (d, J = 1.0 Hz, 1H, H-6), 6.27 (app t, J = 7.1 Hz, 1H, H-1'), 4.51-4.49 (m, 1H, H-3'), 4.09-4.07 (m, 1H, H-4'), 3.94-3.91 (m, 2H, H-5' and H-5''), 3.17-3.10 (m, 2H, -C$H_{2\alpha}$$CH_2SO_3H$), 2.97 (t, J = 6.6 Hz, 2H, -CH_2C$H_{2\beta}$$SO_3H$), 2.33-2.27 (m, 2H, H-2' and H-2''), 1.85 (d, J = 1.0 Hz, 3H, CH_3-Thy); ^{13}C NMR (75 MHz, D_2O) δ = 166.2 (C-4), 151.4 (C-2), 137.1 (C-6), 111.4 (C-5), 85.3 (d, $^{3}J_{C,P}$ = 8.9 Hz, C-4'), 84.7 (C-1'), 70.9 (C-3'), 63.8 (d, $^{2}J_{C,P}$ = 4.9 Hz, C-5'), 52.0 (d, $^{3}J_{C,P}$ = 6.3 Hz, -CH_2C$_\beta$$H_2SO_3H$), 38.4 (C-2'), 36.7 (-C$_\alpha$$H_2CH_2SO_3H$), 11.3 (C$H_3$-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 8.0; HRMS for $C_{12}H_{20}N_3O_{10}PS$ $[M-H]^-$ calcd.: 428.0534, found: 428.0531.

Thymidine-5'-(L-cysteic acid) phosphoramidate ammonium salt (4.2)

The ammonium salt of compound **4.2** was obtained as an off-white solid (0.086 g, 30%) following the general procedure **(A)**, starting from TMP disodium-salt (0.2 g, 0.55 mmol), L-cysteic acid (0.31 g, 1.64 mmol), triethylamine (0.76 mL, 5.46 mmol), DCC (0.34g, 1.64 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (14 mL) at 90 ºC for 3 h. ^{1}H NMR (300 MHz, D_2O) δ = 7.72 (d, J = 1.0 Hz, 1H, H-6), 6.27 (app t, J = 7.0 Hz, 1H, H-1'), 4.51-4.49 (m, 1H, H-3'), 4.09-4.07 (m, 1H, H-4'), 3.96-3.92 (m, 2H, H-5' and H-5''), 3.87-3.83 (m, 1H, C$H$$CO_2H$), 3.16 (d, J = 5.8 Hz, 2H, C$H_2$$SO_3H$), 2.31-2.25 (m, 2H, H-2' and H-2''), 1.86 (d, J = 1.0 Hz, 3H, CH_3-Thy); ^{13}C NMR (75 MHz, D_2O) δ = 178.6 (CO_2H), 166.3 (C-4), 151.5 (C-2), 137.1 (C-6), 111.5 (C-5), 85.5 (d, $^{3}J_{C,P}$ = 9.3 Hz, C-4'), 84.6 (C-1'), 70.9 (C-3'), 63.7 (d, $^{2}J_{C,P}$ = 4.8 Hz, C-5'), 54.5 (d, $^{3}J_{C,P}$ = 6.6 Hz, -CH_2SO_3H), 53.8 (d, $^{2}J_{C,P}$ = 6.1 Hz, C$H$$CO_2H$), 38.2 (C-2'), 11.3 (C$H_3$-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 6.1; HRMS for $C_{13}H_{20}N_3O_{12}PS$ $[M-H]^-$ calcd.: 472.0432, found: 472.0436.

Thymidine-5'-(D-cysteic acid) phosphoramidate ammonium salt (4.3)

The ammonium salt of compound **4.3** was obtained as an off-white solid (0.072 g, 25%) following the general procedure **(A)**, starting from TMP disodium-salt (0.2 g, 0.55 mmol), D-cysteic acid (0.31 g, 1.64 mmol), triethylamine (0.76 mL, 5.46 mmol), DCC (0.34g, 1.64 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (14 mL) at 90 ºC for 3 h. ^{1}H NMR (300 MHz, D_2O) δ = 7.69 (d, J = 1.0 Hz, 1H, H-6), 6.26 (app t, J = 7.1 Hz, 1H, H-1'), 4.52-4.49 (m, 1H, H-3'), 4.09-4.07 (m, 1H, H-4'), 3.96-3.92 (m, 2H, H-5' and H-5''), 3.86-3.82 (m, 1H, C*H*CO_2H), 3.16 (d, J = 5.8 Hz, 2H, CH_2SO_3H), 2.31-2.25 (m, 2H, H-2' and H-2''), 1.86 (d, J = 1.0 Hz, 3H, CH_3-Thy); ^{13}C NMR (75 MHz, D_2O) δ = 178.0 (d, $^3J_{C,P}$ = 5.1 Hz, *C*O_2H), 166.3 (C-4), 151.5 (C-2), 137.2 (C-6), 111.4 (C-5), 85.4 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 84.7 (C-1'), 71.0 (C-3'), 63.9 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 54.5 (d, $^3J_{C,P}$ = 6.6 Hz, *C*H_2SO_3H), 53.5 (d, $^2J_{C,P}$ = 6.6 Hz, *C*HCO_2H), 38.2 (C-2'), 11.3 (*C*H_3-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 6.1; HRMS for HRMS for $C_{13}H_{20}N_3O_{12}PS$ $[M-H]^-$ calcd.: 472.0432, found: 472.0430.

2'-Deoxyadenine-5'-taurine phosphoramidate ammonium salt (4.12)[11]

The ammonium salt of compound **4.12** was obtained as a white solid (0.45 g, 71%) following the general procedure (A), starting from dAMP (0.441 mg, 1.333 mmol), taurine (0.5 g, 4.00 mmol), triethylamine (1.48 mL, 10.66 mmol), DCC (0.82 g, 4.00 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (40 mL) at 90 ºC for 3 h. Spectral data are in accordance with literature report. [Ref. 11]

2'-Deoxyadenine-5'-(L-cysteic acid) phosphoramidate ammonium salt (4.13)[11]

The ammonium salt of compound **4.13** was obtained as a white solid (0.23 g, 32%) following the general procedure **(A)**, starting from dAMP (441.5 mg, 1.333 mmol), L-cysteic acid (338 mg, 1.999 mmol), triethylamine (1.48 mL, 10.66 mmol), DCC (0.82 g, 4.00 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (40 mL) at 90 ºC for 3 h. Spectral data are in accordance with literature report. [Ref. 11]

Thymidine-5'-*O*-(*N*-methyltaurine) phosphoramidate triethylammonium salt (4.4)

The triethylammonium salt of compound **4.4** was obtained as a white solid (0.39 g, 74%) following the general procedure **(A)**, starting from TMP-triethylammonium salt (0.43 g, 0.82 mmol), *N*-methyltaurine (0.57 g, 4.09 mmol), triethylamine (0.91 mL, 6.55 mmol) and DCC (1.01 g, 4.91 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (12 mL) at 90 ºC for 5.5 h.^{1}H NMR (500 MHz, D_2O) δ = 7.73 (s, 1H, H-6), 6.20 (app t, J = 7.0 Hz, 1H, H-1'), 4.53-4.52 (m, 1H, H-3'), 4.12-4.11 (m, 1H, H-4'), 3.94-3.91 (m, 2H, H-5' and H-5''), 3.33-3.27 (m, 2H, C$H_{2\alpha}$$CH_2SO_3H$), 3.08 (t, J = 7.0 Hz, 2H, CH_2C$H_{2\beta}$$SO_3H$), 2.58 (d, $^3J_{H,P}$ = 9.5 Hz , 3H, N-CH_3), 2.34-2.31 (m, 2H, H-2' and H-2''), 1.89 (s, 3H, CH_3-Thy); ^{13}C NMR (125 MHz, D_2O) δ = 166.0 (C-4), 151.1 (C-2), 136.7 (C-6), 111.1 (C-5), 84.9 (d, $^3J_{C,P}$ = 9.1 Hz, C-4'), 84.3 (C-1'), 70.6 (C-3'), 63.5 (d, $^2J_{C,P}$ = 5.3 Hz, C-5'), 48.8 (d, $^3J_{C,P}$ = 2.9 Hz, $CH_2C_\beta H_2SO_3H$), 44.5 (d, $^2J_{C,P}$ = 4.3 Hz, $C_\alpha H_2CH_2SO_3H$), 38.0 (C-2'), 33.0 (d, $^2J_{C,P}$ = 3.5 Hz, N-*C*H_3), 11.1 (*C*H_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 8.9; HRMS for $C_{13}H_{22}N_3O_{10}PS$ $[M-H]^-$ calcd.: 442.0691, found: 442.0689.

Thymidine-5'-(3-aminopropanesulfonic acid) phosphoramidate triethylammonium salt (4.5)

The triethylammonium salt of compound **4.5** was obtained as a white solid (0.272 g , 77%) following the general procedure (A), starting from TMP-disodium salt (0.2 g, 0.55 mmol), 3-aminoethanesulfonic acid (0.23 g, 1.64 mmol), triethylamine (0.46 mL, 3.28 mmol) and DCC (0.45 g, 2.18 mmol) at 90 ºC for 4 h.^{1}H NMR (500 MHz, D_2O) δ = 7.76 (s, 1H, H-6), 6.32 (app t, J = 7.0 Hz, 1H, H-1'), 4.55-4.53 (m, 1H, H-3'), 4.13 (br s, 1H, H-4'), 3.99-3.92 (m, 2H, H-5' and H-5''), 22.91-2.86 (m, 4H, -$CH_{2\alpha}$$CH_2CH_2SO_3H$ and -$CH_2CH_2CH_{2\gamma}SO_3H$), 2.39-2.28 (m, 2H, H-2'

and H-2''), 1.89 (s, 3H, *CH*$_3$-Thy), 1.87-1.81 (m, 2H, -$CH_2CH_{2\beta}CH_2SO_3H$); ^{13}C NMR (125 MHz, D_2O) δ = 166.4 (C-4), 151.5 (C-2), 136.9 (C-6), 111.1 (C-5), 85.0 (d, $^3J_{C,P}$ = 9.1 Hz, C-4'), 84.3 (C-1'), 70.7 (C-3'), 63.4 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 48.2 (-$CH_2CH_2C_\gamma H_2SO_3H$), 39.4 (-$C_\alpha H_2CH_2CH_2SO_3H$), 38.1 (C-2'), 26.0 (d, $^3J_{C,P}$ = 6.8 Hz, -$CH_2C_\beta H_2CH_2SO_3H$), 11.2 (*CH*$_3$-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 8.5 ; HRMS for $C_{13}H_{22}N_3O_{10}PS$ [M-H]$^-$ calcd.: 442.0690, found: 442.0668.

***N,N*-2-Diaminoethanesulfonate-disodium salt (4.14)**

To a stirred solution of 2-aminoethane sulfonic acid (2.0 g, 16.0 mmol) and $NaHCO_3$ (1.34 g, 16.0 mmol) in water, was added an aqueous solution (25%) of sodium vinylsulfonate (6.72 mL, 15.2 mmol). The resulting mixture was heated to reflux for 3 d. The reaction mixture was then cooled and the solvent was removed by distillation under vacuum. The remaining solid was purified by column chromatography on silica gel (gradient: IPA:H_2O 50:1, v/v; 20:1, v/v; 10:1, v/v) to provide compound **4.14** (3.58 g, 85%) as a fluffy white solid. ^{1}H NMR (300 MHz, D_2O) δ = 3.56 (t, *J* = 6.6 Hz, 2H, $CH_{2\alpha}CH_2SO_3H$), 3.32 (t, *J* = 6.6 Hz, 2H, $CH_2CH_{2\beta}SO_3H$); ^{13}C NMR (75 MHz, D_2O) δ = 46.2 ($CH_2C_\beta H_2SO_3H$), 43.3 ($C_\alpha H_2CH_2SO_3H$); HRMS for $C_4H_{11}NO_6S_2$ [M-H]$^-$ calcd.: 231.9955, found: 231.9954.

Thymidine-5'-*(N,N*-2-diaminoethanesulfonate) phosphoramidate triethylammonium salt (4.6)

The triethylammonium salt of compound **4.6** (0.243 g, 53%) was obtained as a white solid following the general procedure (**A**), starting from TMP-triethylammonium salt (0.286 g, 0.55 mmol), compound **4.14** (0.76 g, 2.73 mmol), triethylamine (0.91 mL, 6.55 mmol) and DCC (0.68 g, 3.28 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (9 mL) at 90 ºC for 5.5 h. ^{1}H NMR (500 MHz, D_2O) δ = 7.80 (s, 1H, H-6), 6.37 (app t, *J* = 7.0 Hz, 1H, H-1'), 4.61-4.58 (m, 1H, H-3'), 4.18 (br s, 1H, H-4'), 4.05-3.99 (m, 2H, H-5' and H-5''), 3.43-3.36 (m, 4H, $CH_{2\alpha}CH_2SO_3H$), 3.17 (t, *J* = 8.0 Hz, 4H, $CH_2CH_{2\beta}SO_3H$), 2.40-2.37 (m, 2H, H-2' and H-2''), 1.96 (s, 3H, *CH*$_3$-Thy); ^{13}C NMR (125 MHz, D_2O) δ = 166.0 (C-4), 151.2 (C-2), 136.7 (C-6), 111.3 (C-5), 84.9 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 84.4 (C-1'), 70.6 (C-3'), 63.5 (d, $^2J_{C,P}$ = 5.4 Hz, C-5'), 49.5 ($CH_2C_\beta H_2SO_3H$), 41.7 (d, $^2J_{C,P}$ = 4.9 Hz, $C_\alpha H_2CH_2SO_3H$), 38.1 (C-2'), 11.2 (*C*H$_3$-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 7.4 ; HRMS for $C_{14}H_{24}N_3O_{13}PS_2$ [M-H]$^-$ calcd.: 536.0415, found: 536.0418.

General procedure for taurine-amino acid synthons synthesis

DCC (1.1 eq.) was added to a solution of *N*-Boc-protected amino acid (1 eq.) and *N*-hydroxy succinimide (1.04 eq.) in dry THF (~4 mL/mmol) at 0 ºC, and the mixture was stirred for an additional 0.5 h at same temperature, then for 1.5 h at room temperature. The resulting DCU was filtered off, and the filtrate was concentrated *in vacuo* to afford the activated ester. The crude residue was then dissolved in dioxane (3 mL/mmol), and to this solution was added a solution of taurine sodium salt (1 eq.) in water (2 mL/mmol) at room temperature. The reaction mixture was stirred at room temperature for 14 h, followed by evaporation of the organic phase *in vacuo*. The aqueous layer was washed with ethyl acetate (3 x) and was then loaded on a column packed with IR-120B (H^+ form) resin. The aqueous eluent was evaporated to dryness *in vacuo* to give a crystalline powder, which was recrystallized from ethanol-water to give pure H_2N-aminoacid-Taurine-OH as a white crystalline solid.

H-Gly-Tau-OH (4.16a)

Compound **4.16a** was obtained as a white solid (0.95 g, 91%) according to the general procedure, starting from DCC (1.29 g, 6.27 mmol), Boc-Gly-OH **4.15a** (1.0 g, 5.70 mmol), *N*-hydroxy

succinimide (0.68 g, 5.93 mmol) in THF (20 mL) and taurine sodium salt (0.84 g, 5.70 mmol) in dioxane-water (1:1, 30 mL). ^{1}H NMR (300 MHz, D_2O) δ = 3.82 (s, 2H, CH_2-Gly), 3.67 (t, J = 6.0 Hz, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.13 (t, J = 6.8 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H); ^{13}C NMR (75 MHz, D_2O) δ = 166.5 (CO-Gly), 49.1 (CH$_2C_\beta$H$_2$SO$_3$H), 40.1 (CH$_2$-Gly), 34.8 (C_αH$_2$CH$_2$SO$_3$H); HRMS for $C_4H_{10}N_2O_4S$ [M-H]$^-$ calcd.: 181.0288, found: 181.0290.

H-His-Tau-OH (4.16b)

Compound **4.16b** was obtained as a white solid (1.55 g, 52%) according to the general procedure, starting from DCC (2.58 g, 12.54 mmol), Boc-His-OH **4.15b** (2.92 g, 11.40 mmol), *N*-hydroxy succinimide (1.36 g, 11.86 mmol) in THF (44 mL) and taurine sodium salt (1.68 g, 11.40 mmol) in dioxane-water (1:1, 60 mL). ^{1}H NMR (300 MHz, D_2O) δ = 8.05 (s, 1H, His Ar-H), 7.17 (s, 1H, His Ar-H), 4.13 (t, J = 6.9 Hz, 1H, His-CH), 3.62-3.54 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.18 (d, J = 6.9 Hz, 2H, His-CH_2), 3.08-2.98 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H); ^{13}C NMR (75 MHz, D_2O) δ = 169.6 (CO-His), 135.4 (Ar C-His), 129.6 (Ar C-His), 116.9 (Ar C-His), 52.9 (αC-His), 49.0 (CH$_2C_\beta$H$_2$SO$_3$H), 34.8 (C_αH$_2$CH$_2$SO$_3$H), 28.0 (βC-His); HRMS for $C_8H_{14}N_4O_4S$ [M-H]$^-$ calcd.: 261.0663, found: 261.0663.

Thymidine-5'-*N*-(Gly-Tau-OH)-phosphoramidate ammonium salt (4.7)

The ammonium salt of compound **4.7** was obtained as a white solid (0.29 g , 68%) following the general procedure **(A)**, starting from TMP-disodium salt (0.3 g, 0.82 mmol), H-Gly-Tau-OH (0.49 g, 2.70 mmol), triethylamine (0.57 mL, 4.09 mmol) and DCC (0.78 g, 3.78 mmol) in a 4:1 ^{t}BuOH/H$_2$O mixture (8 mL) at 90 ºC for 4.5 h. ^{1}H NMR (500 MHz, DMSO-d_6) δ = 8.19 (t, J = 5.4 Hz, 1H, NH-Gly), 7.82 (d, J = 1.0 Hz, 1H, H-6), 6.20 (dd, J = 7.8, 6.0 Hz, 1H, H-1'), 4.27-4.25 (m, 1H, H-3'), 3.86-3.65 (m, 1H, H-4'), 3.76-3.74 (m, 2H, H-5' and H-5''), 3.68 (br s, 1H, 3'-OH), 3.41-3.37 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.21 (d, J = 6.3 Hz, 2H, CH_2-Gly), 3.62 (t, J = 6.3 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.14-2.00 (m, 2H, H-2' and H-2''), 1.80 (d, J = 1 Hz, 3H, CH_3-Thy); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ = 171.7 (d, $^3J_{C,P}$ = 5.1 Hz, CO-Gly), 163.9 (C-4), 150.6 (C-2), 136.4 (C-6), 109.9 (C-5), 86.2 (d, $^3J_{C,P}$ = 7.8 Hz, C-4'), 83.8 (C-1'), 71.2 (C-3'), 63.9 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 50.3 (CH$_2C_\beta$H$_2$SO$_3$H), 45.8 (CH$_2$-Gly), 39.2 (C-2'), 35.1 (C_αH$_2$CH$_2$SO$_3$H), 12.2 (CH$_3$-Thy); ^{31}P NMR (202 MHz, DMSO-d_6) δ = 5.1 ; HRMS for $C_{14}H_{23}N_4O_{11}PS$ [M-H]$^-$ calcd.: 485.0749, found: 485.0750.

Thymidine-5'-(*N*-His-Tau-OH)-phosphoramidate triethylammonium salt (4.8)

The triethylammonium salt of compound **4.8** was obtained as white solid (0.26 g, 41%) following the general procedure **(A)**, starting from TMP-disodium salt (0.3 g, 0.82 mmol), H-His-Tau-OH (0.97 g, 3.68 mmol), triethylamine (0.57 mL, 4.09 mmol) and DCC (0.85 g, 4.09 mmol) in a 4:1 ^{t}BuOH/H$_2$O mixture (10 mL) at 90 ºC for 4.5 h. ^{1}H NMR (500 MHz, D_2O) δ = 7.61 (d, J =1 Hz, 1H, H-6), 7.59 (d, J = 0.5 Hz, 1H, imidazole), 6.88 (s, 1H, imidazole), 6.32 (app t, J = 6.5 Hz, 1H, H-1'), 4.43-4.41 (m, 1H, H-3'), 4.04 (br s, 1H, H-4'), 3.81-3.75 (m, 3H, H-5', H-5'' and αCH-His), 3.54-3.47 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 2.93-2.91 (m, 2H, βCH_2-His), 2.90-2.88 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.28-2.18 (m, 2H, H-2' and H-2''), 1.84 (d, J = 1Hz, 3H, CH$_3$); ^{13}C NMR (125 MHz, D_2O) δ = 177.0 (CO-His), 173.9 (C-4), 157.0 (C-2), 137.9 (C-6), 136.9 (Ar C-His), 113.0 (C-5), 86.4 (d, $^3J_{C,P}$ =9.3 Hz, C-4'), 86.1 (C-1'), 72.5 (C-3'), 65.2 (d, $^2J_{C,P}$ = 4.2 Hz, C-5'), 57.1 (αCH-His), 50.6 (CH$_2C_\beta$H$_2$SO$_3$H), 40.0 (C-2'), 35.9 (C_αH$_2$CH$_2$SO$_3$H), 32.2 (βCH$_2$-His), 13.5 (CH$_3$); ^{31}P NMR (202 MHz, D_2O) δ = 5.3; HRMS for $C_{18}H_{27}N_6O_{11}PS$ [M-H]$^-$ calcd.: 565.1123, found: 565.1126.

_N_-Cbz-Asp(OBn)-Tau-OH (4.17)

Compound **4.17** was prepared according to a modification of the general procedure for taurine-amino acid synthon synthesis (ion-exchange chromatography was not required) starting from *N*-Cbz-Asp(OBn)-OH (2.5 g, 6.99 mmol), *N*-hydroxy succinimide (0.84 g, 7.27 mmol), DCC (1.59 g, 7.70 mmol) in THF (30 mL), and taurine (0.92 g, 7.35 mmol) and $NaHCO_3$ (1.17 g, 13.99 mmol) in dioxane-water (1:1, 60 mL). The crude product was resuspended in water, cooled in an ice-bath and acidified to pH = 5 with the addition of 1N HCl. The aqueous layer was then lyophilized and the crude residue was purified by column chromatography on silica gel (gradient: IPA:H_2O 100:0, v/v; 30:1, v/v; 20:1, v/v) to give compound **4.17** (2.89 g, 89%) as a white solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.06 (t, J = 5.2 Hz, 1H, N*H*-Tau), 7.67 (d, J = 8.3 Hz, 1H, N*H*-Asp), 7.37-7.30 (m, 10H, 2 x ArH of OBn), 5.08 and 5.02 (2 s, 2H, 2 x OCH_2Ph), 4.40-4.33 (m, 1H, αC*H*-Asp), 3.37-3.33 (m, 2H, C$H_{2\alpha}$$CH_2SO_3H$), 2.87-2.60 (m, 2H, βC$H_2$-Asp), 2.56 (t, J = 7.3 Hz, 2H, CH_2C$H_{2\beta}$$SO_3H$); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 170.2 (βCO-Asp), 169.8 (αCO-Asp), 155.8 (OCONH), 136.9 (1C of OCH_2Ph), 136.1 (1C of OCH_2Ph), 128.4 (Ar-C), 128.3 (Ar-C), 127.9 (Ar-C), 127.8 (Ar-C), 127.7 (Ar-C), 65.6 (2 x OCH_2Ph), 51.5 (αC-Asp), 50.1 ($CH_2C_\beta H_2SO_3H$), 36.4 (βC-Asp), 35.7 ($C_\alpha H_2CH_2SO_3H$); HRMS for $C_{21}H_{24}N_2O_8S$ [M-H]$^-$ calcd.: 463.1180, found: 463.1181.

H-Asp-Tau-OH (4.18)

To a stirring solution of **4.17** (2.89 g, 2.44 mmol) in EtOH:H_2O (5:1, 80 mL) was added 10% Pd/C (0.5 g, 15% w/w) and the mixture was hydrogenated at atmospheric pressure using a balloon filled with H_2 for 16 h. The catalyst was then removed by filtration over a pad of Celite and the filtrate was concentrated under reduced pressure to give compound **4.18** as a white solid (1.49 g, quan.). The resulting solid was analysed for purity and used in the next step without any further purification. ^{1}H NMR (300 MHz, D_2O) δ = 4.22-4.18 (m, 1H, αC*H*-Asp), 3.66-3.60 (m, 2H, C$H_{2\alpha}$$CH_2SO_3H$), 3.11 (t, J = 6.7 Hz, 2H, CH_2C$H_{2\beta}$$SO_3H$), 2.84-2.64 (m, 2H, βC$H_2$-Asp); ^{13}C NMR (75 MHz, D_2O) δ = 177.7 (CO_2H), 170.9 (αCO-Asp), 52.5 (αC-Asp), 51.1 ($CH_2C_\beta H_2SO_3H$), 38.5 (βC-Asp), 37.0 ($C_\alpha H_2CH_2SO_3H$); HRMS for $C_6H_{12}N_2O_6S$ [M-H]$^-$ calcd.: 239.0343, found: 239.0339.

General procedure for phosphoramidites synthesis starting from 2'-deoxynucleoside-5'-phosphorimidazolide sodium salts (B)

To a stirred solution of 2'-deoxynucleoside-5'-phosphorimidazolide sodium salt (1 eq.) and amino-acid (2.5-3.0 eq.) in DMF, Et_3N (5-10 eq.) was added and the mixture was stirred vigorously until total dissolution. The reaction was stirred at 35 °C for 7 days and then quenched by addition of water. The solvent was removed under reduced pressure and the resulting crude material was purified by column chromatography on silica gel (gradient: IPA/H_2O/Et_3N 20:1.5:0.5, v/v/v; 15:1.5:0.5, v/v/v; 10:1.5:0.5, v/v/v) to provide the corresponding phosphoramidate as a trietylammonium salt which was further purified by RP-HPLC (50 mmol TEAB in 98% H_2O + 2% ACN and 50 mmol TEAB in 50% H_2O + 50% ACN) to give the pure product as a white solid. The isolated product was freeze-dried repeatedly until constant mass.

Thymidine-5'-(Asp-Tau-OH) phosphoramidate triethylammonium salt (4.9)

The triethylammonium salt of compound **4.9** was obtained as a white solid (0.17 g, 53%) following the general procedure **(B)**, starting from thymidine 5'-phosphorimidazolide-sodium salt (0.15 g, 0.38 mmol), compound **4.18** (0.229 g, 0.951 mmol) and Et_3N (0.41 mL, 3.04 mmol) in DMF (4 mL). ^{1}H NMR (500 MHz, D_2O) δ = 7.81 (s, 1H, H-6), 6.39 (app t, J = 7.0 Hz, 1H, H-1'), 4.58-4.57 (m, 1H, H-3'), 4.20 (br s, 1H, H-4'), 4.03-4.01 (m, 2H, H-5', H-5''), 3.84-3.79 (m, 1H,

αC*H*-Asp), 3.64-3.60 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.13 (t, *J* = 7.0 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.71 (dd, *J* = 16.0, 5.4 Hz, 1H, βC*H'*-Asp), 2.60 (dd, *J* = 15.7, 5.1 Hz, 1H, βC*H''*-Asp), 2.40-2.37 (m, 2H, H-2' and H-2''), 1.95 (s, 3H, CH_3-Thy); ^{13}C NMR (125 MHz, D$_2$O) δ = 177.7 (CO$_2$H), 175.6 (d, $^3J_{C,P}$ = 6.6 Hz, αCO-Asp), 166.0 (C-4), 151.2 (C-2), 136.8 (C-6), 111.1 (C-5), 85.1 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 84.4 (C-1'), 70.8 (C-3'), 63.6 (d, $^2J_{C,P}$ = 4.6 Hz, C-5'), 52.6 (αCH-Asp), 49.0 (CH$_2$C$_\beta$H$_2$SO$_3$H), 39.9 (d, $^3J_{C,P}$ = 3.5 Hz, βCH$_2$-Asp), 38.1 (C-2'), 34.5 (C$_\alpha$H$_2$CH$_2$SO$_3$H), 11.1 (CH$_3$-Thy); ^{31}P NMR (202 MHz, D$_2$O) δ = 5.8; HRMS for C$_{16}$H$_{25}$N$_4$O$_{13}$PS [M-H]$^-$ calcd.: 543.0803, found: 543.0813.

***N*-(Ethyl acetate)-2-aminoethanesulfonic acid (4.19)**

To a stirred solution of 2-aminoethane sulfonic acid (4.0 g, 31.96 mmol) in a 1:1 dioxane/H$_2$O mixture (80 mL), NaHCO$_3$ (2.68 g, 31.96 mmol) was added, followed by ethyl bromo acetate (3.75 mL, 35.16 mmol). The reaction mixture was then heated at 70 ºC for 24 h. The organic layer was removed under reduced pressure and the remaining aqueous layer was washed with ethyl acetate (2 x 100 mL) and neutralized with 1N HCl. The aqueous layer was then concentrated under reduced pressure and the residue was purified by column chromatography on silica gel (gradient: DCM/MeOH 9:1, v/v; 8:2, v/v; 7.5:2.5, v/v) to afford compound **4.19** (3.24 g, 48%) as a white solid. ^{1}H NMR (300 MHz, D$_2$O) δ = 4.32 (q, *J* = 7.1 Hz , 2H, CO$_2$CH_2CH$_3$), 4.05 (s, 2H, CH_2CO$_2$Et), 3.54 (t, *J* = 6.8 Hz, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.33 (t, *J* = 6.8 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 1.31 (t, *J* = 7.1 Hz, 3H, CO$_2$CH$_2$CH_3); ^{13}C NMR (75 MHz, D$_2$O) δ = 167.8, 64.1, 48.3, 47.1, 43.9, 13.8; HRMS for C$_6$H$_{13}$NO$_5$S [M-H]$^-$ calcd.: 210.04416, found: 210.0439.

***N*-(Ethyl propionate)-2-aminoethanesulfonic acid (4.20)**

To a stirred solution of 2-aminoethane sulfonic acid (4.0 g, 31.96 mmol) in a 1:1 EtOH/H$_2$O mixture (80 mL), NaHCO$_3$ (2.68 g, 31.96 mmol) was added, followed by ethyl acrylate (3.75 mL, 35.16 mmol). The reaction mixture was then stirred at room temperature for 7 days. Ethanol was removed under reduced pressure and the aqueous layer was washed with ethyl acetate (2 x 100 mL) and neutralized with 1N HCl. The aqueous layer was then concentrated under reduced pressure and the residue was purified by column chromatography on silica gel (gradient: DCM/MeOH 9:1, v/v; 8:2, v/v; 7:3, v/v) to afford compound **4.20** (4.17 g, 58%) as a white solid. ^{1}H NMR (300 MHz, D$_2$O) δ = 4.15 (q, *J* = 7.1 Hz , 2H, CO$_2$CH_2CH$_3$), 3.45 (t, *J* = 6.6 Hz, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.35 (t, *J* = 6.5 Hz, 2H, C$H_{2\alpha}$CH$_2$CO$_2$Et), 3.23 (t, *J* = 6.6 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.80 (t, *J* = 6.5 Hz, 2H, CH$_2$C$H_{2\beta}$CO$_2$Et), 1.20 (t, *J* = 7.1 Hz, 3H, CO$_2$CH$_2$CH_3); ^{13}C NMR (75 MHz, D$_2$O) δ = 172.1, 62.1, 46.1, 43.1, 42.9, 29.9, 12.9; HRMS for C$_7$H$_{15}$NO$_5$S [M-H]$^-$ calcd.: 224.0598, found: 224.0598.

Thymidine-5'-*O*-[*N*-(ethylacetate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.21)

The triethylammonium salt of compound **4.21** was obtained as a white solid (123 mg, 52%) following the general procedure **(B)**, starting from thymidine-5'-phosphorimidazolide sodium salt (130 mg, 0.33 mmol), compound **4.19** (195 mg, 0.92 mmol), triethylamine (0.28 mL, 1.98 mmol) and DMF (4 mL). ^{1}H NMR (300 MHz, D$_2$O) δ = 7.69 (s, 1H, H-6), 6.37 (t, *J* = 7.1 Hz, 1H, H-1'), 4.58-4.54 (m, 1H, H-3'), 4.17-4.19 (m, 3H, H-4' and OCH_2CH$_3$), 4.05-3.98 (m, 2H, H-5' and H-5''), 3.85 (d, *J* = 10.6 Hz, 2H, CH_2CO$_2$Et), 3.48-3.37 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.18-3.10 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H) (merged with Et$_3$N), 2.37-2.32 (m, 2H, H-2' and H-2''), 1.92 (s, 3H, CH$_3$), 1.27-1.22 (m, 3H, OCH$_2$CH_3) (merged with Et$_3$N); ^{13}C NMR (75 MHz, D$_2$O) δ = 173.5 (d, $^3J_{C,P}$ = 2.4 Hz, CO$_2$Et), 171.0 (C-4), 155.1 (C-2), 136.5 (C-6), 111.5 (C-5), 85.0 (d, $^3J_{C,P}$ = 9.4 Hz, C-4'), 84.6 (C-1'), 71.0 (C-3'), 63.9 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 61.6 (O*C*H$_2$CH$_3$), 49.7 (CH$_2$C$_\beta$*H*$_2$SO$_3$H),

48.5 (d, $^{2}J_{C,P}$ = 5.5 Hz, CH_2CO$_2$Et), 43.4 (d, $^{2}J_{C,P}$ = 4.7 Hz, C$_\alpha H_2$CH$_2$SO$_3$H), 38.5 (C-2'), 13.0 (OCH$_2$CH_3), 11.9 (CH$_3$); ^{31}P NMR (121 MHz, D$_2$O) δ = 7.1; HRMS for $C_{16}H_{26}N_3O_{12}PS$ [M-H]- calcd.: 514.0902, found: 514.0902.

2'-Deoxyadenosine-5'-*O*-[*N*-(ethylacetate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.22)

The triethylammonium salt of compound **4.22** was obtained as a white solid (47 mg, 38%) following the general procedure **(B)**, starting from 2'-deoxyadenosine-5'-phosphorimidazolide sodium salt (70 mg, 0.17 mmol), compound **4.19** (103 mg, 0.49 mmol), triethylamine (0.145 mL, 1.04 mmol) and DMF (3 mL). ^{1}H NMR (300 MHz, D$_2$O) δ = 8.40 (s, 1H, H-8), 8.13 (s, 1H, H-2), 6.40 (t, J = 6.7 Hz, 1H, H-1'), 4.72-4.68 (m, 1H, H-3'), 4.24-4.23 (m, 1H, H-4'), 4.05-3.95 (m, 4H, H-5',H-5'' and OCH_2CH$_3$), 3.70 (d, J = 10.5 Hz, 2H, CH_2CO$_2$Et), 3.41-3.28 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.11-3.03 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.91-2.74 (m, 1H, H-2'), 2.64-2.53 (m, 1H, H-2''), 1.13 (t, J = 7.1 Hz, 3H, OCH$_2$CH_3); ^{13}C NMR (75 MHz, D$_2$O) δ = 173.3 (d, $^{3}J_{C,P}$ = 2.8 Hz, CO$_2$Et), 155.0 (C-6), 152.2 (C-2), 148.1 (C-4), 139.4 (C-8), 118.1 (C-5), 85.7 (d, $^{3}J_{C,P}$ = 9.3 Hz, C-4'), 83.3 (C-1'), 71.1 (C-3'), 63.8 (d, $^{2}J_{C,P}$ = 5.2 Hz, C-5'), 61.5 (OCH$_2$CH$_3$), 49.7 (CH$_2C_\beta$H$_2$SO$_3$H), 48.4 (d, $^{2}J_{C,P}$ = 5.3 Hz, CH$_2$CO$_2$Et), 43.4 (d, $^{2}J_{C,P}$ = 4.6 Hz, C_αH$_2$CH$_2$SO$_3$H), 38.6 (C-2'), 12.3 (OCH$_2C$H$_3$); ^{31}P NMR (121 MHz, D$_2$O) δ = 7.2; HRMS for $C_{16}H_{25}N_6O_{10}PS$ [M-H]$^-$ calcd.: 523.1017, found: 523.1019.

2'-Deoxyguanosine-5'-*O*-[*N*-(ethyl acetate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.23)

The triethylammonium salt of compound **4.23** was obtained as an off-white solid (58 mg, 22%) following the general procedure **(B)**, starting from 2'-deoxyguanosine-5'-phosphorimidazolide sodium salt (150 mg, 0.36 mmol), compound **4.19** (212 mg, 1.00 mmol), triethylamine (0.30 mL, 2.15 mmol) and DMF (4 mL). ^{1}H NMR (300 MHz, D$_2$O) δ = 7.99 (s, 1H, H-8), 6.26 (app t, J = 7.0 Hz, 1H, H-1'), 4.65-4.63 (m, 1H, H-3'), 4.17-4.15 (m, 1H, H-4'), 4.04 (q, J = 7.0 Hz, 2H, OCH_2CH$_3$), 3.96-3.93 (m, 2H, H-5' and H-5''), 3.74 (d, J = 10.5 Hz, 2H, CH_2CO$_2$Et), 3.40-3.33 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.01-2.96 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.83-2.71 (m, 1H, H-2'), 2.50-2.43 (m, 1H, H-2''), 1.13 (t, J = 7.4 Hz, 3H, OCH$_2$CH_3); ^{13}C NMR (75 MHz, D$_2$O) δ = 173.4 (d, $^{3}J_{C,P}$ = 2.9 Hz, CO$_2$Et), 159.4 (C-6), 153.4 (C-2), 151.0 (C-4), 135.4 (C-8), 117.0 (C-5), 85.3 (d, $^{3}J_{C,P}$ = 9.3 Hz, C-4'), 82.6 (C-1'), 71.2 (C-3'), 64.0 (d, $^{2}J_{C,P}$ = 5.4 Hz, C-5'), 61.6 (OCH$_2$CH$_3$), 49.8 (CH$_2C_\beta$H$_2$SO$_3$H), 48.5 (d, $^{2}J_{C,P}$ = 5.4 Hz, CH$_2$CO$_2$Et), 43.4 (d, $^{2}J_{C,P}$ = 4.6 Hz, C_αH$_2$CH$_2$SO$_3$H), 38.4 (C-2'), 12.9 (OCH$_2C$H$_3$); ^{31}P NMR (121 MHz, D$_2$O) δ = 7.2; HRMS for $C_{16}H_{25}N_6O_{11}PS$ [M-H]$^-$ calcd.: 539.0967, found: 539.0969.

2'-Deoxycytidine-5'-*O*-[*N*-(ethyl acetate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.24)

The triethylammonium salt of compound **4.24** was obtained as a white solid (130 mg, 47%) following the general procedure **(B)**, starting from 2'-deoxycytidine-5'-phosphorimidazolide sodium salt (150 mg, 0.39 mmol), compound **4.19** (234 mg, 1.11 mmol), triethylamine (0.33 mL, 2.37 mmol) and DMF (4 mL).^{1}H NMR (300 MHz, D$_2$O) δ = 7.94 (d, J = 7.6 Hz, 1H, H-6), 6.29 (app t, J = 6.8 Hz, 1H, H-1'), 6.08 (d, J = 7.5 Hz, 1H, H-5), 4.53-4.49 (m, 1H, H-3'), 4.17-4.10 (m, 3H, H-4' and OCH_2CH$_3$), 4.03-3.98 (m, 2H, H-5' and H-5''), 3.82 (d, J = 10.7 Hz, 2H, CH_2CO$_2$Et), 3.44-3.35 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.19-3.09 (unresolved m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H, merged with Et$_3$N), 2.43-2.21 (m, 2H, H-2' and H-2''), 1.27-1.22 (unresolved m, 3H, OCH$_2$CH_3, merged with Et$_3$N); ^{13}C NMR (75 MHz, D$_2$O) δ = 173.4 (d, $^{3}J_{C,P}$ = 2.5 Hz, CO$_2$Et), 165.8 (C-4), 157.1 (C-2), 141.2 (C-6), 96.1 (C-5), 85.6 (C-1'), 85.4 (d, $^{3}J_{C,P}$ = 9.3 Hz, C-4'), 70.7 (C-3'), 63.7

(d, $^{2}J_{C,P}$ = 5.2 Hz, C-5'), 61.6 (O*C*H_2CH$_3$), 49.7 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 48.6 (d, $^{2}J_{C,P}$ = 5.5 Hz, *C*H$_2$CO$_2$Et), 43.4 (d, $^{2}J_{C,P}$ = 4.7 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 39.3 (C-2'), 13.0 (OCH$_2$*C*H$_3$); ^{31}P NMR (121 MHz, D$_2$O) δ = 7.2; HRMS for C$_{15}$H$_{25}$N$_4$O$_{11}$PS [M-H]$^-$ calcd.: 499.0905, found: 499.0898.

Thymidine-5'-*O*-[*N*-(ethyl propionate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.25)

The triethylammonium salt of compound **4.25** was obtained as a white solid (0.204 g, 51%) following the general procedure **(A)**, starting from TMP-triethylammonium salt (0.286 g, 0.55 mmol), compound **4.20** (0.553 g, 2.46 mmol), DCC (0.563 g, 2.73 mmol), triethylamine (0.38 mL, 2.73 mmol) in a 4:1 ^{t}BuOH/H$_2$O mixture (11 mL) at 90 °C for 6 h. ^{1}H NMR (500 MHz, D$_2$O) δ = 7.75 (s, 1H, H-6), 6.35 (app t, *J* = 7.1 Hz, 1H, H-1'), 4.58-4.56 (m, 1H, H-3'), 4.13-4.12 (m, 1H, H-4'), 4.09-4.03 (m, 2H, OC*H*$_2$CH$_3$), 3.96-3.90 (m, 2H, H-5' and H-5''), 3.42-3.35 (m, 2H, C*H*$_{2\alpha}$CH$_2$SO$_3$H), 3.34-3.10 (m, 2H, C*H*$_{2\alpha}$CH$_2$CO$_2$Et), 3.12-3.07 (m, 2H, CH$_2$C*H*$_{2\beta}$SO$_3$H), 2.63-2.51 (m, 2H, CH$_2$C*H*$_{2\beta}$CO$_2$Et), 2.41-2.30 (m, 2H, H-2' and H-2''), 1.90 (s, 3H, CH$_3$), 1.19 (t, *J* = 7.1 Hz, 3H, OCH$_2$C*H*$_3$); ^{13}C NMR (125 MHz, D$_2$O) δ = 174.2 (CO$_2$Et), 165.9 (C-4), 151.2 (C-2), 136.8 (C-6), 111.2 (C-5), 84.9 (d, $^{3}J_{C,P}$ = 9.4 Hz, C-4'), 84.2 (C-1'), 70.5 (C-3'), 63.3 (d, $^{2}J_{C,P}$ = 5.0 Hz, C-5'), 61.1 (O*C*H$_2$CH$_3$), 49.5 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 41.7 (d, $^{2}J_{C,P}$ = 4.2 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 41.5 (d, $^{2}J_{C,P}$ = 4.2 Hz, *C*$_\alpha$H$_2$CH$_2$CO$_2$Et), 37.9 (C-2'), 33.2 (CH$_2$*C*$_\beta$H$_2$CO$_2$Et), 12.6 (OCH$_2$*C*H$_3$), 11.1 (CH$_3$); ^{31}P NMR (202 MHz, D$_2$O) δ = 7.7; HRMS for C$_{17}$H$_{28}$N$_3$O$_{12}$PS [M-H]- calcd.: 528.1058, found: 528.1058.

2'-Deoxyadenosine-5'-*O*-[*N*-(ethyl propionate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.26)

The triethylammonium salt of compound **4.26** was obtained as a white solid (237 mg, 53%) following the general procedure **(A)**, starting from dAMP (200 mg, 0.60 mmol), compound **4.20** (611.4 mg, 2.72 mmol), DCC (623.1 mg, 3.02 mmol), triethylamine (0.42 ml, 3.02 mmol) in a 4:1 ^{t}BuOH/H$_2$O mixture (12 mL) at 90 °C for 6 h. ^{1}H NMR (300 MHz, D$_2$O) δ = 8.42 (s, 1H, H-8), 8.21 (s, 1H, H-2), 6.46 (app t, *J* = 6.6 Hz, 1H, H-1'), 4.74-4.69 (m, 1H, H-3'), 4.23-4.22 (m, 1H, H-4'), 3.98-3.94 (m, 2H, OC*H*$_2$CH$_3$), 3.93-3.89 (m, 2H, H-5' and H-5''), 3.32-3.23 (m, 2H, C*H*$_{2\alpha}$CH$_2$SO$_3$H), 3.18-3.12 (unresolved m, 2H, C*H*$_{2\alpha}$CH$_2$CO$_2$Et merged with Et$_3$N), 3.06-2.99 (m, 2H, CH$_2$C*H*$_{2\beta}$SO$_3$H), 2.90-2.81 (m, 1H, H-2'), 2.64-2.55 (m, 1H, H-2''), 2.42-2.27 (m, 2H, CH$_2$C*H*$_{2\beta}$CO$_2$Et), 1.09 (t, *J* = 7.1 Hz, 3H, OCH$_2$C*H*$_3$); ^{13}C NMR (75 MHz, D$_2$O) δ = 174.2 (CO$_2$Et), 155.2 (C-6), 152.4 (C-2), 148.4 (C-4), 139.5 (C-8), 118.3 (C-5), 85.6 (d, $^{3}J_{C,P}$ = 9.3 Hz, C-4'), 83.4 (C-1'), 70.8 (C-3'), 63.5 (d, $^{2}J_{C,P}$ = 5.0 Hz, C-5'), 61.2 (O*C*H$_2$CH$_3$), 49.9 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 42.1 (d, $^{2}J_{C,P}$ = 4.7 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 41.9 (d, $^{2}J_{C,P}$ = 4.6 Hz, *C*$_\alpha$H$_2$CH$_2$CO$_2$Et), 38.5 (C-2'), 33.5 (CH$_2$*C*$_\beta$H$_2$CO$_2$Et), 12.8 (OCH$_2$*C*H$_3$); ^{31}P NMR (121 MHz, D$_2$O) δ = 8.1; HRMS for C$_{17}$H$_{27}$N$_6$O$_{10}$PS [M-H]- calcd.: 537.1174, found: 537.1168.

General deprotection procedure for 2'-deoxynucleoside-5'-*O*-[*N*-(alkyl acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salts (4.10-4.11 and 4.27-4.30) synthesis (C)

A solution of 2'-deoxynucleoside-5'-*O*-[*N*-(ethyl alkylate)-2-aminoethane sulfonic acid] phosphoramidate trietylammonium salt (1 mmol) in a 0.4M NaOH solution in MeOH/H$_2$O (4:1v/v, 15 mL) was stirred at room temperature for 3 h and then 1M TEAB buffer (8 mL) was added. All the volatiles were removed first by rotary evaporation (bath temp 10 °C) and then by lyophilization. The crude residue was purified by RP-HPLC (50 mmol TEAB in 98% H$_2$O + 2% ACN and 50 mmol TEAB in 50% H$_2$O + 50% ACN) to give pure 2'-deoxynucleoside-5'-*O*-[*N*,*N*-

(alkyl acid)-2-aminoethane sulfonic acid] phos-phoramidate trietylammonium salt after the product was freeze-dried repeatedly to constant mass.

Thymidine-5'-*O*-[*N*-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.10)

The triethylammonium salt of compound **4.10** was obtained as a white solid (18.5 mg, 84%) following the general procedure (**C**), starting from compound 21 (20 mg, 0.028 mmol), 0.4M NaOH in MeOH/H_2O (4:1v/v, 0.42 mL) and 1M TEAB (0.22 mL). ^{1}H NMR (600 MHz, D_2O) δ = 7.78 (s, 1H, H-6), 6.37 (app t, J = 7.1 Hz, 1H, H-1'), 4.54-4.52 (m, 1H, H-3'), 4.14-4.13 (m, 1H, H-4'), 4.02-4.01 (m, 2H, H-5' and H-5''), 3.57 (d, J = 10.6 Hz, 2H, CH_2CO$_2$H), 3.39-3.35 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.13-3.10 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.38-2.28 (m, 2H, H-2' and H-2''), 1.91 (s, 3H, CH$_3$); ^{13}C NMR (150 MHz, D2O) δ = 178.6 (d, $^3J_{C,P}$ = 3.9 Hz, CO$_2$H), 166.4 (C-4), 151.6 (C-2), 137.1 (C-6), 111.6 (C-5), 85.5 (d, $^3J_{C,P}$ = 9.4 Hz, C-4'), 84.7 (C-1'), 71.1 (C-3'), 63.9 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 49.7 (d, $^2J_{C,P}$ = 4.0 Hz, CH$_2$CO$_2$H), 49.1 (CH$_2$C$_\beta$H$_2$SO$_3$H), 43.1 (d, $^2J_{C,P}$ = 3.2 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 38.4 (C-2'), 11.4 (-CH3); ^{31}P NMR (202 MHz, D_2O) δ = 7.8; HRMS for $C_{16}H_{26}N_3O_{12}PS$ [M-H]- calcd.: 486.0589, found: 486.0591.

2'-Deoxyadenosine-5'-*O*-[*N*-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.27)

The triethylammonium salt of compound **4.27** was obtained as a white solid (41 mg, 81%) following the general procedure (C), starting from compound **4.22** (46 mg, 0.063 mmol), 0.4M NaOH in MeOH/H_2O (4:1v/v, 1.0 mL) and 1M TEAB (0.5 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.41 (s, 1H, H-8), 8.09 (s, 1H, H-2), 6.39 (t, J = 6.8 Hz, 1H, H-1'), 4.66-4.63 (m, 1H, H-3'), 4.21-4.20 (m, 1H, H-4'), 4.01-3.92 (m, 2H, H-5' and H-5''), 3.50 (d, J = 9.0 Hz, 2H, CH_2CO$_2$H), 3.35-3.30 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.03-3.00 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.80-2.75 (m, 1H, H-2'), 2.57-2.52 (m, 1H, H-2''); ^{13}C NMR (125 MHz, D_2O) δ = 178.2 (d, $^3J_{C,P}$ = 4.7 Hz, CO$_2$H), 154.6 (C-6), 151.7 (C-2), 147.8 (C-4), 139.1 (C-8), 117.7 (C-5), 85.4 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 82.9 (C-1'), 70.6 (C-3'), 63.3 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 49.3 (d, $^2J_{C,P}$ = 3.7 Hz, CH$_2$CO$_2$H), 48.7 (CH$_2$C$_\beta$H$_2$SO$_3$H), 42.7 (d, $^2J_{C,P}$ = 4.0 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 38.3 (C-2'); ^{31}P NMR (202 MHz, D_2O) δ = 8.1; HRMS for $C_{14}H_{21}N_6O_{10}PS$ [M-H]$^-$ calcd.: 495.0705, found: 495.0703.

2'-Deoxyguanosine-5'-*O*-[*N*-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.28)

The triethylammonium salt of compound **4.28** was obtained as an off-white solid (42 mg, 67%) following the general procedure **(C)**, starting from compound **4.23** (58 mg, 0.078 mmol), 0.4M NaOH in MeOH/H_2O (4:1v/v, 1.17 mL) and 1M TEAB (0.6 mL). ^{1}H NMR (500 MHz, D_2O) δ = 7.78 (s, 1H, H-8), 6.25 (app t, J = 6.9 Hz, 1H, H-1'), 4.64-4.62 (m, 1H, H-3'), 4.18-4.16 (m, 1H, H-4'), 4.00-3.95 (m, 2H, H-5' and H-5''), 3.52 (d, J = 8.8 Hz, 2H, CH_2CO$_2$H), 3.37-3.31 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.07-3.04 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.80-2.75 (m, 1H, H-2'), 2.48-2.43 (m, 1H, H-2''), 1.91 (s, 3H, CH$_3$); ^{13}C NMR (125 MHz, D2O) δ = 178.3 (d, $^3J_{C,P}$ = 4.8 Hz, CO$_2$H), 158.3 (C-6), 153.2 (C-2), 150.8 (C-4), 136.9 (C-8), 115.5 (C-5), 85.3 (d, $^3J_{C,P}$ = 9.0 Hz, C-4'), 82.9 (C-1'), 70.8 (C-3'), 63.5 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 49.3 (d, $^2J_{C,P}$ = 3.8 Hz, CH$_2$CO$_2$H), 48.7 (CH$_2$C$_\beta$H$_2$SO$_3$H), 42.7 (d, $^2J_{C,P}$ = 4.4 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 37.9 (C-2'); ^{31}P NMR (202 MHz, D_2O) δ = 8.1; HRMS for $C_{14}H_{21}N_6O_{11}PS$ [M-H]- calcd.: 511.0654, found: 511.0651.

2'-Deoxycytidine-5'-*O*-[*N*-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.29)

The triethylammonium salt of compound **4.29** was obtained as a white solid (112 mg, 78%)

following the general procedure (**C**), starting from compound **4.24** (130 mg, 0.185 mmol), 0.4M NaOH in MeOH/H_2O (4:1v/v, 2.77 mL) and 1M TEAB (1.4 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.00 (d, *J* = 7.5 Hz, 1H, H-6), 6.37 (app t, *J* = 6.8 Hz, 1H, H-1'), 6.16 (d, *J* = 7.5 Hz, 1H, H-5), 4.58-4.55 (m, 1H, H-3'), 4.21-4.20 (m, 1H, H-4'), 4.09-4.07 (m, 2H, H-5' and H-5''), 3.60 (dd, *J* = 8.9, 1.8 Hz, 2H, CH_2CO$_2$H), 3.46-3.40 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.19-3.15 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.46-2.41 (m, 1H, H-2'), 2.34-2.29 (m, 1H, H-2''); ^{13}C NMR (125 MHz, D_2O) δ = 178.3 (d, $^3J_{C,P}$ = 4.9 Hz, CO$_2$H), 165.5 (C-4), 157.0 (C-2), 140.9 (C-6), 96.0 (C-5), 85.3 (C-1'), 85.2 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 70.4 (C-3'), 63.3 (d, $^2J_{C,P}$ = 5.0 Hz, C-5'), 49.5 (d, $^2J_{C,P}$ = 3.5 Hz, *C*H$_2$CO$_2$H), 48.8 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 42.8 (d, $^2J_{C,P}$ = 4.3 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 39.0 (C-2'); ^{31}P NMR (202 MHz, D_2O) δ = 8.1; HRMS for $C_{13}H_{21}N_4O_{11}PS$ [M-H]- calcd.: 471.0592, found: 471.0594.

Thymidine-5'-*O*-[*N*-(propionic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.11)

The triethylammonium salt of compound **4.11** was obtained as a white solid (164 mg, 88%) following the general procedure (**C**), starting from compound **4.25** (170 mg, 0.23 mmol), 0.4M NaOH in MeOH/H_2O (4:1v/v, 3.48 mL) and 1M TEAB (1.85 mL). ^{1}H NMR (500 MHz, DMSO-*d*$_6$) δ = 7.78 (d, *J* = 1.2 Hz, 1H, H-6), 6.20 (dd, *J* = 7.9, 6.0 Hz, 1H, H-1'), 4.27-4.25 (m, 1H, H-3'), 3.86-3.85 (m, 1H, H-4'), 3.74-3.72 (m, 2H, H-5' and H-5''), 3.17-3.13 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.10-3.05 (m, 2H, C$H_{2\alpha}$CH$_2$CO$_2$H), 2.60-2.57 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.41-2.38 (m, 2H, CH$_2$C$H_{2\beta}$CO$_2$H), 2.14-2.01 (m, 2H, H-2' and H-2''), 1.81 (d, *J* = 1.2 Hz, 3H, CH$_3$); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ = 173.7 (CO$_2$H), 163.9 (C-4), 150.6 (C-2), 136.2 (C-6), 109.9 (C-5), 86.1 (d, $^3J_{C,P}$ = 7.6 Hz, C-4'), 83.8 (C-1'), 71.1 (C-3'), 64.0 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 51.2 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 44.0 (d, $^2J_{C,P}$ = 3.0 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 43.0 (d, $^2J_{C,P}$ = 3.3 Hz, *C*$_\alpha$H$_2$CH$_2$CO$_2$H), 39.6 (C-2', merged with DMSO), 35.7 (CH$_2$*C*$_\beta$H$_2$CO$_2$H), 12.1 (CH3); ^{31}P NMR (202 MHz, DMSO-d$_6$) δ = 6.3; HRMS for $C_{16}H_{26}N_3O_{12}PS$ [M-H]- calcd.: 500.0745, found: 500.0746.

2'-Deoxyadenosine-5'-*O*-[*N*-(propionic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (4.30)

The triethylammonium salt of compound **4.30** was obtained as a white solid (142 mg, 86%) following the general procedure (**C**), starting from compound **4.26** (150 mg, 0.202 mmol), 0.4M NaOH in MeOH/H_2O (4:1v/v, 3.04 mL) and 1M TEAB (1.50 mL). ^{1}H NMR (500 MHz, D_2O) δ = 8.38 (s, 1H, H-8), 8.10 (s, 1H, H-2), 6.39 (app t, *J* = 6.6 Hz, 1H, H-1'), 4.69-4.67 (m, 1H, H-3'), 4.20 (br s, 1H, H-4'), 3.91-3.89 (m, 2H, H-5' and H-5''), 3.30-3.24 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.17-3.07 (m, 2H, C$H_{2\alpha}$CH$_2$CO$_2$H, merged with Et$_3$N), 3.03-2.99 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.82-2.76 (m, 1H, H-2'), 2.59-2.54 (m, 1H, H-2''), 2.31-2.28 (t, 2H, *J* = 7.9 Hz, CH$_2$C$H_{2\beta}$CO$_2$H); ^{13}C NMR (125 MHz, D_2O) δ = 180.2 (CO$_2$H), 154.7 (C-6), 151.8 (C-2), 147.7 (C-4), 139.0 (C-8), 117.8 (C-5), 85.3 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 83.0 (C-1'), 70.4 (C-3'), 63.1 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 49.5 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 43.1 (d, $^2J_{C,P}$ = 4.2 Hz, *C*$_\alpha$H$_2$CH$_2$CO$_2$H), 41.2 (d, $^2J_{C,P}$ = 4.6 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 38.3 (C-2'), 36.8 (CH$_2$*C*$_\beta$H$_2$CO$_2$H); ^{31}P NMR (202 MHz, D_2O) δ = 8.3; HRMS for $C_{15}H_{23}N_6O_{10}PS$ [M-H]- calcd.: 509.0861, found: 509.0853.

Thymidine-5'-diphosphate triethylammonium salt (4.32)

To a stirred suspension of 5'-*O*-tosyl-thymidine **4.31** (0.20 g , 0.50 mmol) in dry acetonitrile (0.5 mL) was added tris(terta-n-butylammonium) hydrogen phosphate (0.682 g, 0.75 mmol). The resulting solution was stirred at room temperature for 36 h. Upon reaction completion, the mixture was concentrated *in vacuo*, diluted with water (5 mL) and lyophilized. The crude product was purified by Source 15Q ion-exchance column chromatography from a gradient of 0-100%

1M TEAB solution. The product was freeze-dried repeatedly to constant mass, yielding **4.32** as a white triethylammonium salt solid (0.239 g, 67%). ^{1}H NMR (300 MHz, D_2O) δ = 7.75 (d, J = 1.0 Hz, 1H, H-6), 6.35 (app t, J = 7.0 Hz, 1H, H-1'), 4.65-4.61 (m, 1H, H-3'), 4.18-4.14 (m, 3H, H-4', H-5' and H-5''), 2.39-2.34 (m, 2H, H-2' and H-2''), 1.92 (d, J = 1.0 Hz, 3H, CH_3); ^{13}C NMR (75 MHz, D_2O) δ = 166.2 (C-4), 151.4 (C-2), 137.0 (C-6), 111.4 (C-5), 85.1 (d, $^{3}J_{C,P}$ = 9.1 Hz, C-4'), 84.6 (C-1'), 70.7 (C-3'), 65.0 (d, $^{2}J_{C,P}$ = 5.5 Hz, C-5'), 38.2 (C-2'), 11.3 (CH_3); ^{31}P NMR (121 MHz, D_2O) δ = -10.9 (d, $J_{P,P}$ = 20.0 Hz, β-P), -11.5 (d, $J_{P,P}$ = 20.0 Hz, α-P) HRMS for $C_{10}H_{16}N_2O_{11}P_2$ [M-H]$^-$ calcd.: 401.0156, found: 401.0155.

Thymidine-5'-[β-*N*-(2-aminoethanesulfonic acid)] diphosphoramidate triethylammonium salt (4.33)

The triethylammonium salt of compound **4.33** was obtained as a white solid (95 mg, 69%) following the general procedure (**A**), starting from compound **4.32** (0.12 g , 0.17 mmol), taurine (0.085 g, 0.68 mmol), triethylamine (0.12 mL, 0.85 mmol) and DCC (0.175 g, 0.85 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (4 mL) at 85 ºC for 4.5 h.^{1}H NMR (500 MHz, D_2O) δ = 7.76 (s, 1H, H-6), 6.32 (app t, J = 7.0 Hz, 1H, H-1'), 4.61-4.59 (m, 1H, H-3'), 4.15-4.14 (m, 1H, H-4'), 4.13-4.12 (m, 2H, H-5' and H-5''), 3.28-3.23 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.07 (t, J = 7.0 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.38-2.27 (m, 2H, H-2' and H-2''), 1.89 (s, 3H, CH_3); ^{13}C NMR (125 MHz, D_2O) δ = 166.0 (C-4), 151.1 (C-2), 136.8 (C-6), 111.1 (C-5), 84.8 (d, $^{3}J_{C,P}$ = 10.0 Hz, C-4'), 84.2 (C-1'), 70.4 (C-3'), 64.6 (d, $^{2}J_{C,P}$ = 5.0 Hz, C-5'), 51.8 (d, $^{3}J_{C,P}$ = 7.5 Hz, CH$_2$C$_\beta$H$_2$SO$_3$H), 37.9 (C-2'), 36.6 (C_αH$_2$CH$_2$SO$_3$H), 11.1 (CH_3); ^{31}P NMR (202 MHz, D_2O) δ = -2.6 (d, $J_{P,P}$ = 22.0 Hz, β-P), -11.7 (d, $J_{P,P}$ = 22.0 Hz, α-P); HRMS for $C_{12}H_{21}N_3O_{13}P_2S$ [M-H]$^-$ calcd.: 508.0197, found: 508.0198.

Thymidine-5'-[γ-*N*-(2-aminoethanesulfonic acid)] triphosphoramidate triethylammonium salt (4.35)

DCC (0.153 g, 0.744 mmol) was added to a stirred solution of dTTP triethylammonium salt **4.34** (0.11 g, 0.124 mmol) in dry DMF (2 mL) under an argon atmosphere and stirred for 3.5 h at room temperature. The dTTP solution was then added dropwise to a solution of taurine (0.093 g, 0.744 mmol) in a 1:1 miture of dry methanol and dry DMF (16 mL) and the reaction mixture was left to stir for 6 h at room temperture. The reaction was then diluted with water (50 mL) and the precipitate was filtered off. The clear solution was collected and lyophilized. The crude residue was redissolved in 0.01M TEAB buffer and purified by Source-15 Q ion-exchange resin column chromatography. The products adhering to the column were eluted with a linear buffer gradient of 0 to 1 M TEAB. The desired product was isolated by using 0.6–0.8M TEAB and it was then further purified by HPLC to give compound **4.35** as a white triethylammonium salt solid (0.022 g, 18%).^{1}H NMR (500 MHz, D_2O) δ = 7.64 (s, 1H, H-6), 6.35 (app t, J = 7.0 Hz, 1H, H-1'), 4.62-4.60 (m, 1H, H-3'), 4.18-4.11 (m, 3H, H-4', H-5' and H-5''), 3.29-3.24 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.00-2.96 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.34-2.26 (m, 2H, H-2' and H-2''), 1.85 (s, 3H, CH_3); ^{13}C NMR (125 MHz, D_2O) δ = 173.1 (C-4), 156.5 (C-2), 136.0 (C-6), 111.4 (C-5), 84.5 (d, $^{3}J_{C,P}$ = 9.8 Hz, C-4'), 84.2 (C-1'), 70.3 (C-3'), 64.8 (d, $^{2}J_{C,P}$ = 5.5 Hz, C-5'), 51.9 (d, $^{3}J_{C,P}$ = 7.2 Hz, CH$_2$C$_\beta$H$_2$SO$_3$H), 42.1 (C_αH$_2$CH$_2$SO$_3$H), 37.9 (C-2'), 11.9 (CH_3); ^{31}P NMR (202 MHz, D_2O) δ = -2.7 (d, $J_{P,P}$ = 20.6 Hz, γ-P), -12.0 (d, $J_{P,P}$ = 19.8 Hz, α-P), -23.3 (app t, $J_{P,P}$ = 20.2 Hz, β-P); HRMS for $C_{12}H_{22}N_3O_{16}P_3S$ [M-H]$^-$ calcd.: 587.9861, found: 587.9866.

General Protocol for DNA polymerase reaction

The primer P_1 was purchased from IDT, whilst all the templates T1-5 were purchased from

Eurogentec. Primer oligonucleotides were labeled with 5' [γ-33P]-ATP (Perkin Elmer) using T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's protocol. The labeled primers were further purified using Illustra MicroSpin G-25 Column (GE Healthcare) and then annealed with the corresponding template oligonucleotides in a 1:2 molar ratio by heating the mixture at 75 °C for 5 min, followed by slow cooling to room temperature. The DNA polymerisation mixtures contained 125 nM primer-template complex, 1X reaction buffer (supplied with the enzyme), different concentrations of building blocks (125 µM, 500 µM and 1 mM) and 0.01 $U.\mu l^{-1}$ Therminator, 0.01 $U.\mu l^{-1}$ Vent (-exo) or 0.05 $U.\mu l^{-1}$ Klenow (-exo) polymerases (New England Biolabs). The reaction was performed either at 37 °C (mesophilic polymerase) or at 75 °C (thermophilic polymerases) and aliquots were taken after 15, 30 and 60 min. In the control reaction, 50 µM of the corresponding natural deoxynucleoside triphosphate was used. All the polymerase reactions were quenched by addition of a double volume of gel loading buffer (90% formamide, 50mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol). Samples were heated either at 75 °C (mesophilic polymerase) or at 90 °C (thermophilic polymerases) for 5 min prior to separation on a 0.4mm 20% denaturing polyacrylamide gel. The bands were then visualized using phosphorimaging (Perkin Elmer).

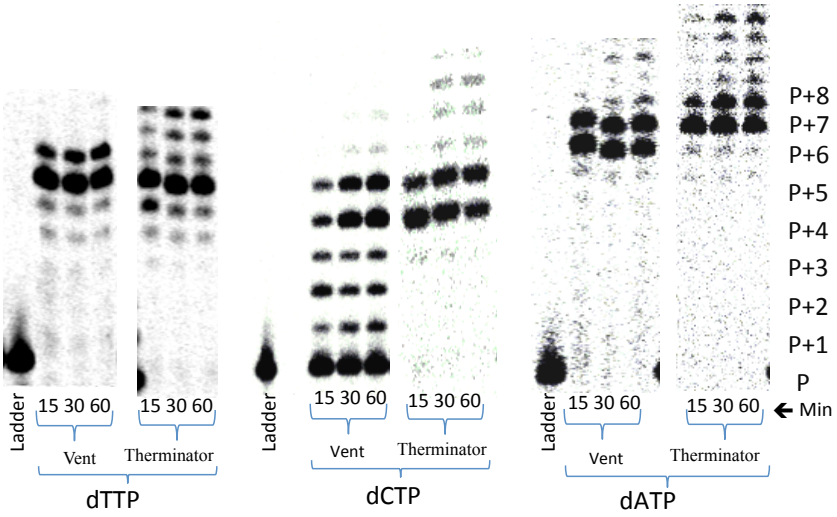


Figure 4-6. Profile of chain elongation of the control **dTTP**, **dCTP**, and **dATP** as substrate into the P1:T1, P1:T3 and P1:T2 respectively by two different DNA polymerase; Vent (0.01 U.µl-1) and Therminator (0.01 U.µl-1).

Steady-state kinetics of single nucleotide incorporation

Primer 5' /5Cy5/CAGGAAACAGCTATGAC 3' was annealed with template 3' GTCCTTTGTCGATACTG**CAAAA** 5' in a 1:2 molar ratio by heating the mixture at 75 °C for 5 min, followed by slow cooling to room temperature. A series of reactions with different enzyme concentrations and different time points were performed to obtain the optimum conditions satisfying the 'single completed hit' principle[21]. The final DNA polymerisation mixtures each contained 125 nM primer-template complex, 1X reaction buffer (supplied with the enzyme), building blocks (for modified building blocks the concentration ranged between 15 µM and 1 mM, whilst for natural nucleoside triphosphates the concentration ranged between 25 nM and 2 µM) and 0.001 $U.\mu l^{-1}$ Klenow fragment (New England Biolabs). The reactions were performed at 37 °C and aliquots were taken after 30 seconds. All polymerase reactions were quenched by

adding a double volume of gel loading buffer (90 % formamide, 50 mM EDTA and 0.05% bromophenol blue) and heated at 90 °C for 5 min. The samples were separated on a 1 mm 15% denaturing polyacrylamide gel and gel bands were visualized using Ettan DIGE Imager (GE Healthcare). The gel bands were then quantified using ImageQuant TL 1D version 7.0 (GE Healthcare) and the kinetics parameters (V_{max} and K_m) were determined by fitting the data to a non-linear Michaelis–Menten regression using GraphPad Prism Software version 5.0.

4.6. REFERENCES

[1] F. H. Westheimer, *Science*, 1987, **235**, 1173-1178.
[2] C. M. Joyce and T. A. Steitz, *Annu. Rev. Biochem.*, 1994, **63**, 777-822.
[3] P. Herdewijn and P. Marliere, *Chem. Biodivers.*, 2009, **6**, 791-808.
[4] P. Herdewijn and P. Marliere, *FEBS Lett.*, 2012, **586**, 2049-2056 and references cited therein.
[5] S. Q. Yang, M. Froeyen, E. Lescrinier, P. Marliere and P. Herdewijn, *Org. Biomol. Chem.*, 2011, **9**, 111-119.
[6] M. A. Kertesz, *Fems Microbiol. Rev.*, 2000, **24**, 135-175.
[7] J. R. van der Ploeg, E. Eichhorn and T. Leisinger, *Arch. Microbiol.*, 2001, **176**, 1-8.
[8] J. R. van der Ploeg, M. A. Weiss, E. Saller, H. Nashimoto, N. Saito, M. A. Kertesz and T. Leisinger, *J. Bacteriol.*, 1996, **178**, 5438-5446.
[9] J. R. van der Ploeg, R. Iwanicka-Nowicka, T. Bykowski, M. M. Hryniewicz and T. Leisinger, *J. Biol. Chem.*, 1999, **274**, 29358-29365.
[10] E. Eichhorn, J. R. vanderPloeg, M. A. Kertesz and T. Leisinger, *J. Biol. Chem.*, 1997, **272**, 23031-23036.
[11] S. Yang, M. Froeyen, E. Lescrinier, P. Marliere and P. Herdewijn, *Org. Biomol. Chem.*, 2011, **9**, 111-119.
[12] X.-P. Song, C. Bouillon, E. Lescrinier and P. Herdewijn, *Chem. Biodivers.*, 2012, **9**, 2685-2700.
[13] K. Ienaga, K. Higashiura, Y. Toyomaki, H. Matsuura and H. Kimura, *Chem. Pharm. Bull. (Tokyo)*, 1988, **36**, 70-77.
[14] A. Giraut, X.-p. Song, M. Froeyen, P. Marliere and P. Herdewijn, *Nucleic Acids Res.*, 2010, **38**, 2541-2550.
[15] X.-P. Song, C. Bouillon, E. Lescrinier and P. Herdewijn, *Chembiochem*, 2011, **12**, 1867-1879.
[16] W. Zhang and Z. Huang, *Org. Lett.*, 2011, **13**, 2000-2003.
[17] V. J. Davisson, D. R. Davis, V. M. Dixit and C. D. Poulter, *J. Org. Chem.*, 1987, **52**, 1794-1801.
[18] S. Martic, M. Labib, D. Freeman and H.-B. Kraatz, *Chem. Eur. J.*, 2011, **17**, 6744-6752.
[19] A. Giraut, R. Abu El-Asrar, P. Marliere, M. Delarue and P. Herdewijn, *Chembiochem*, 2012, **13**, 2439-2444.
[20] B. Singer, F. Chavez, M. F. Goodman, J. M. Essigmann and M. K. Dosanjh, *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 8271-8274.
[21] S. Creighton, Bloom, L. B., Goodman, M. F., *Method. Enzymol*, 1993, 232-256.

Chapter 5

Synthesis of new biocarrier-nucleotide systems for cellular delivery in bacterial auxotrophic strains

This chapter is based on an article published by S. De, E. Groaz, M. Maiti, V. Pezo, P. Marlière, and P. Herdewijn, Tetrahedron ***2014****, 70 (46), 8843–8851.*

ABSTRACT

In search for a delivery approach for thymidine monophosphate (TMP) in bacterial cells, we have synthesized a series of conjugates of TMP with biotin having an oxymethyleneoxy ester, a carboxy ester and different carboxamide linkers between the carboxyl group of biotin and the 3'-OH group of TMP. The synthetic strategy starts from 5'-O-(dibenzylphosphate)-thymidine having the linkers already connected at the 3'-position. Likewise, kanamycin A was linked at the 3'-position of TMP using a carbamoyl or thioethyl carbamoyl group.

5.1. INTRODUCTION

Creating artificial genetic systems capable of supporting Darwinian evolution is of prominent importance to synthetic biology, not only to achieve expansion of the genetic alphabet but specially for the genetic reprogramming of engineered microorganisms.[1] The *in vivo* synthesis of xeno nucleic acids (XNAs), however, is thought to entail high intracellular consumption of unnatural nucleotide building blocks, which will thus need to be provided to the host cell, as most modified nucleosides are poorly phosphorylated by natural kinases. Nucleotides however suffer from low cell-penetrating ability and generally undergo fast enzymatic dephosphorylation before being taken up. Their bacterial uptake is further complicated by the complex multi-layered wall structure, which makes prokaryotic cells commonly more difficult to penetrate compared to eukaryotic membranes.

Thus far, efforts have focused on overexpression of endogenous kinases to broad enzyme-substrate specificity,[2] and more recently, on the transfection of *Escherichia* coli (*E. coli*) with a plasmid encoding an algal nucleotide triphosphate transporter to enhance triphosphates uptake.[3] Despite the groundbreaking value of this last approach, which resulted in the first *in vivo* replication of an unnatural DNA base pair, its universal applicability to all unnatural bases and types of backbone modifications is yet unpredictable.

In the previous chapter, we discussed our interest in the development of a delivery tool for active uptake of nucleotides in bacterial cells based on a nutritional selection approach.[4] The general concept consists in linking the information system to an extracellular nutrient essential for the cell, thus taking advantage of natural transport pathways. A biocleavable covalent bond is designed to maintain the conjugate intact during cellular uptake, but allows facile release of the nucleotide intracellularly. Building on previous studies, thymidine monophosphate is selected as archetypal nucleotide system.[5] The relevant TMP conjugate is supplied within the nutrient medium to a bacterial strain deleted for the ThyA gene encoding for thymidilate synthase, which is known to unconditionally require thymine or thymidine for growth.

Because of the novelty of our delivery approach, we decided to investigate a large diversity of biological carriers, such as peptides and vitamins. We describe here the synthesis of biotin−TMP conjugates. The appeal of this choice is due to the fact that biotin-negative mutants are also well-established auxotrophs and can provide additional functional evidence.[6, 7]

Expanding upon this idea, we also wish to prove the same concept using an inverse selection method. As in most cases, antimicrobial compounds penetrate prokaryotic cell membranes effectively through relatively non-specific transport systems, we describe here also the synthesis of conjugates of nucleotides with kanamycin A. Aminoglycosides (AGs) are clinically relevant drugs, which show broad-spectrum antibacterial activity against both gram-negative and gram-positive species, binding irreversibly to the prokaryotic ribosome and mainly interfering with peptide elongation at the 30S subunit.[8, 9] Although the detailed mechanism by which AGs penetrate into the bacterial cytoplasm

remains unclear, their uptake is known to be a multistep energy requiring process that can be accumulated against a concentration gradient.[10, 11] In this case, the observation of toxicity would ensure uptake and processing of the nucleotide by the cell.

5.2. RESULTS AND DISCUSSION

Our retrosynthetic analysis for the planned carrier−TMP target compounds is shown in Fig. 5-1. It becomes apparent that those conjugates can be accessed by either linking the carrier unit at the 3'-position of thymidine followed by 5'-*O*-phosphorylation (route 5.1), or *viceversa* (route 5.2).

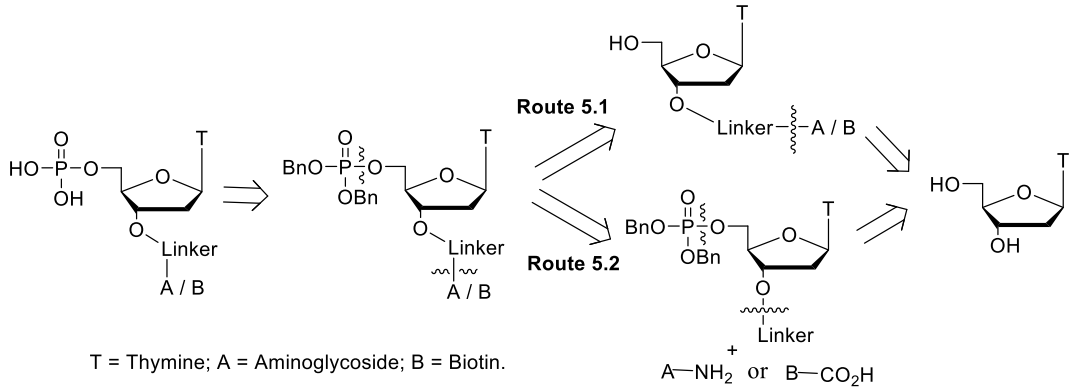
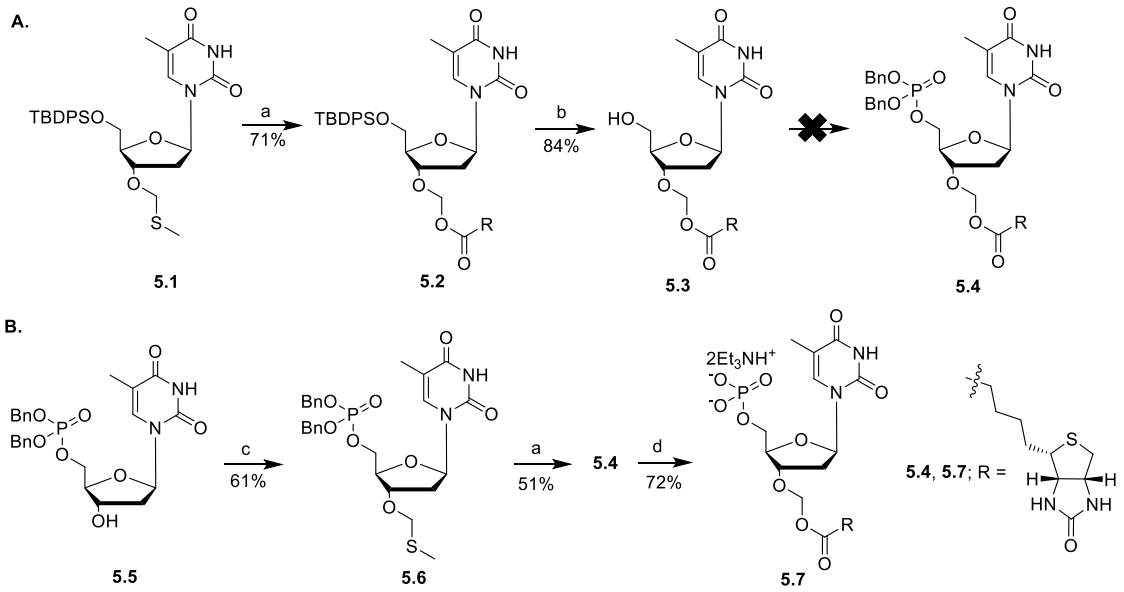


Figure 5-1. Retrosynthetic analysis of carrier-TMP conjugates.

Initially, we decided to follow route 5.1 (Fig. 5-1) by applying the same optimized conditions employed for the synthesis of peptide−TMP conjugates connected through an oxymethyleneoxy ester bond.[4] Coupling of biotin at the 3'-position of methylene thiomethyl thymidine derivative **5.1**, obtained in turn from thymidine in 60% yield over two steps, and subsequent removal of the 5'-silyl group, proceeded efficiently giving compound **5.3** in good yield; however, further phosphorylation caused oxidation at the sulfur atom of biotin, during the conversion step from P^{III} to P^{V}, giving undesired sulfoxide and sulfone by-products (Scheme 5-1A). Various mild conditions were attempted to attain selective phosphorus oxidation, including dilute solutions of I_2,[12] m-chloroperbenzoic acid (MCPBA)[13] and (1*S*)-(+)-(10-camphorsulfonyl) oxaziridine[14] but none gave the desired product. Switching to the alternative disconnection, 5'-O-(dibenzylphosphate)-thymidine **5.5** was therefore prepared in four steps and subsequently subjected to coupling with biotin, leading to the formation of the desired compound **5.7** after deprotection (Scheme 5-1B). However, we found that the Pummerer rearrangement of compound **5.5** to **5.6** was less effective than in the previous route, probably due to side reaction of the electrophilic sulfur intermediate at the nucleophilic phosphate center.

During the last step, formation of side products resulting from conjugate decomposition was observed, pointing at instability of this type of linker under hydrogenation conditions.



Scheme 5-1. Synthesis of 3'-*O*-methyleneacyloxy-TMP biotin conjugate **5.7**. Reagents and conditions: (a) (i) SO_2Cl_2, DCM, 0 °C to rt, 2 h, (ii) biotin, DBU, DCM, rt, 24 h; (b) $Et_3N{\cdot}3HF$, THF, rt, 72 h; (c) DMSO, Ac_2O, AcOH, rt, 48 h; (d) 10% Pd/C (Degussa), Et_3N, MeOH, rt, 16 h.

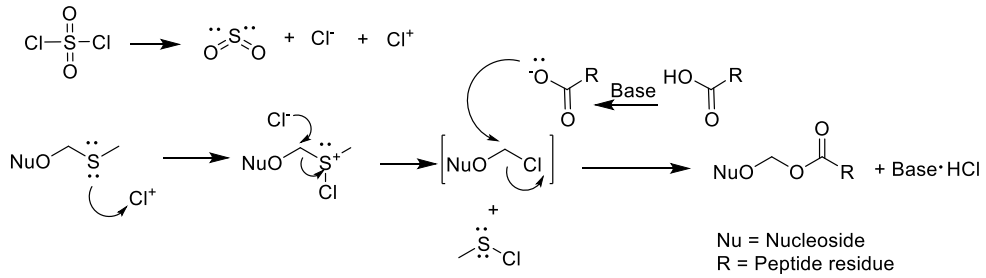
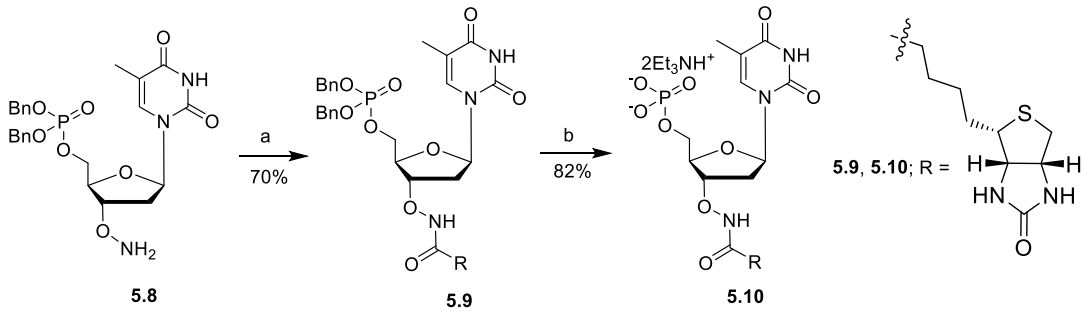


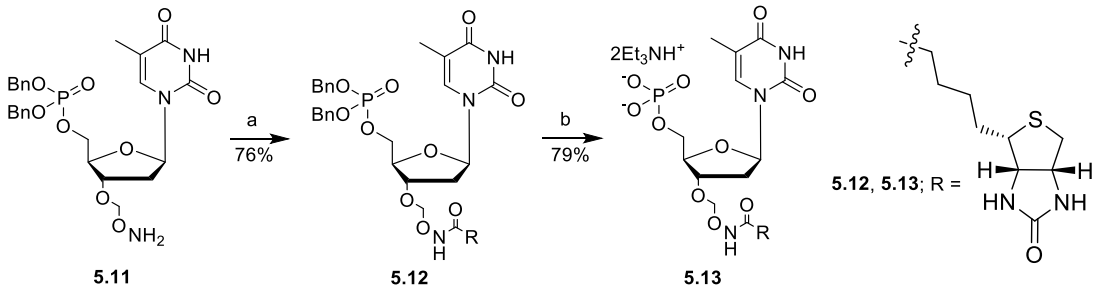
Figure 5-2. Proposed mechanism for the formation of **5.2** and **5.4** via α-halogenation of the corresponding thiomethyl derivative.



Scheme 5-2. Synthesis of 3'-*O*-(carboxamide)-TMP biotin conjugate **5.10**. Reagents and conditions: (a) biotin, DCC, DMAP, DCM/DMF, rt, 24 h; (b) 10% Pd/C (Degussa), MeOH, rt, 24 h.

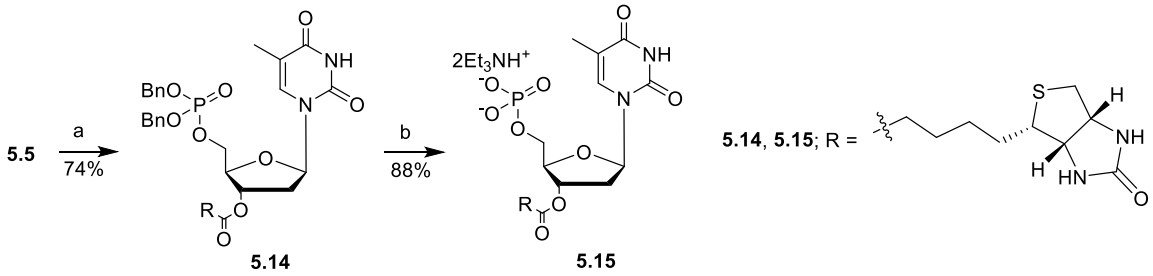
Different biotin−TMP analogues with carboxyamido-type linkers **5.10** (Scheme 5-2) and **5.13** (Scheme 5-3) were prepared adopting previously optimized coupling and debenzylation procedures, which required access respectively to 3'-*O*-amino thymidylate **5.8** (obtained in four steps from thymidine) and 3'-*O*-aminoxymethyl thymidylate **5.11** (obtained in six steps from thymidine).[4] In general good to excellent isolated yields of products were obtained and linker

cleavage did not occur in the last step.



Scheme 5-3. Synthesis of 3'-*O*-methyloxycarboxamide-TMP biotin conjugate **5.6**. Reagents and conditions: (a) biotin, DCC, DMAP, DCM/DMF, rt, 24 h; (b) 10% Pd/C (Degussa), Et_3N, MeOH, rt, 24 h.

The 3'-carboxy ester conjugate **5.15** was also prepared similarly, starting from **5.5** in 65% overall yield over two steps (Scheme 5-4).



Scheme 5-4. Synthesis of 3'-ester-TMP biotin conjugate **5.15**. Reagents and conditions: (a) biotin, DCC, DMAP, DCM/DMF, rt, 24 h; (b) 10% Pd/C (Degussa), MeOH, rt, 24 h.

As route 5.2 (Figure 5-1) proved to be the most profitable for generating new biotin conjugates, we chose a similar strategy to connect TMP to kanamycin A by means of a 3'-carbamoyl linker, which gave access to conjugates **5.22** and **5.24** (Figure 5-3), and also led to a structural reassignment of the aminoglycosidic unit.

As cationic amino functionalities of aminoglycosides are considered to be a crucial structural requirement for enhanced uptake,[15] we sought to retain all four amino groups of rings I–III in kanamycin A. To construct the relevant conjugates efficiently, we exploited the differences in chemical reactivity between the primary hydroxy and amino functionalities present in kanamycin A. We started by replacing the primary hydroxyl group at the 6"-position of **5.16** with a thioethanolamino group, which would serve for the formation of the desired carbamoyl linkage with TMP. The synthesis of compound **5.19** has been reported by Arya et al.,[16] however, we adopted here a shorter synthetic route by modifying a more recent procedure by Tor et. al.[17] (Scheme 5-5). To begin the synthesis, we first needed to selectively protect all four amino groups of kanamycin A. Among the several protecting strategies described in the literature, we opted for the classical Boc protection,

which gave compound **5.17** in excellent yield. Considering the good leaving group properties of triisopropylbenzenesulfonyl chloride (TIBS-Cl), compound **5.17** was selectively sulphonylated at the 6''-primary hydroxyl group to its corresponding sulfonate **5.18,** which upon further treatment with 2-aminoethane thiol hydrochloride in the presence of cesium carbonate, gave compound **5.19**.

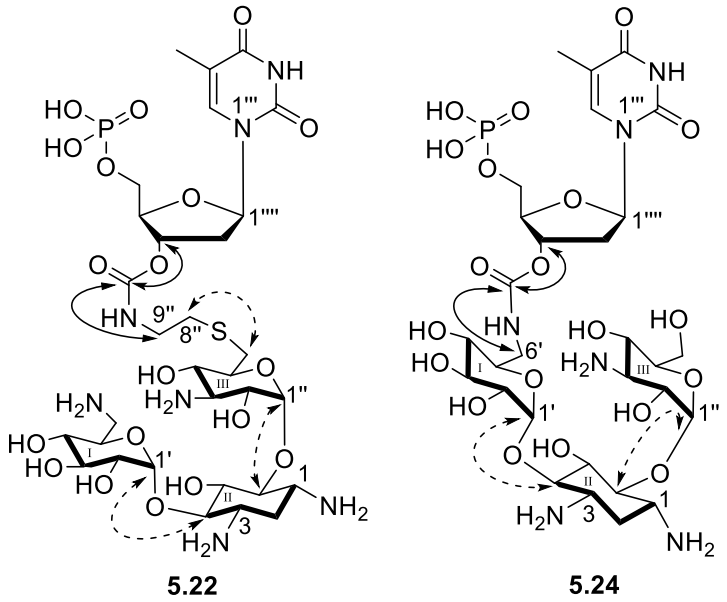
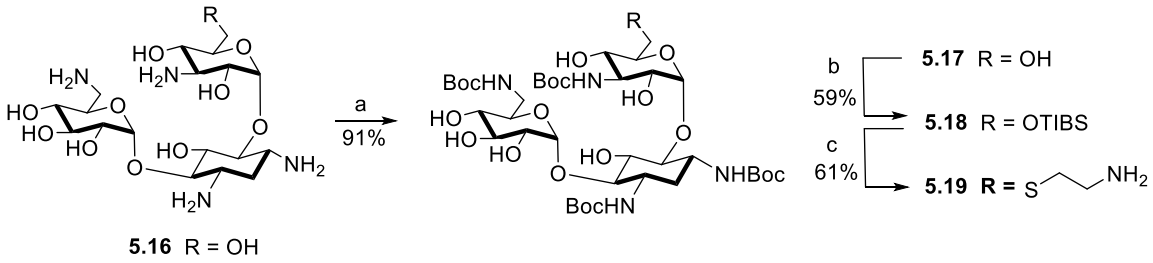


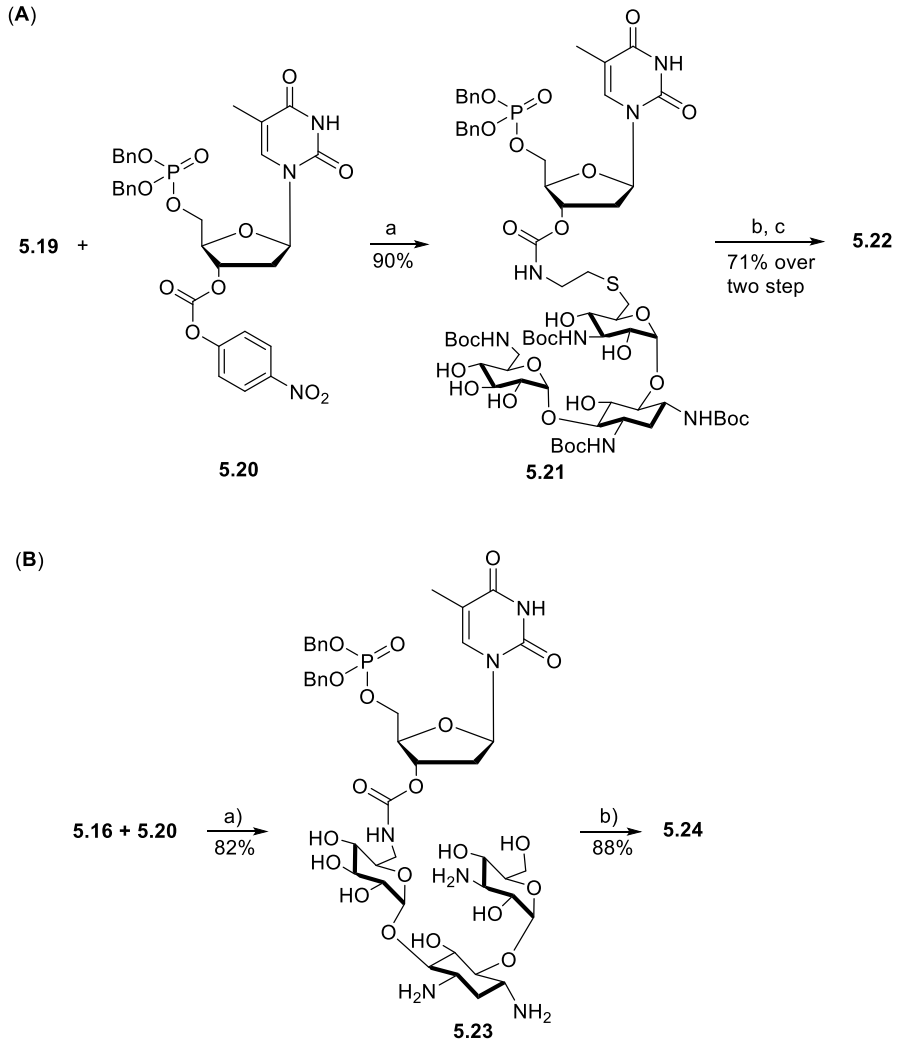
Figure 5-3. Structures of kanamycin A−TMP conjugates **5.22** and **5.24** and key HMBC correlations used for structure determination. Plain arrows (⟷) indicate HMBC correlations with the carbamate (OCONH) linker (cf. Figure 5-4A/B) whilst dotted arrows (⟷) show HMBC correlation used for other peak assignments.



Scheme 5-5. Synthesis of kanamycin A cysteamine **5.19**. Reagents and conditions: (a) Boc_2O, Et_3N, DMF/H_2O, 50 °C, 6 h; (b) TIBSCl, pyridine, rt, 16 h; (c) $HCl{\cdot}NH_2CH_2CH_2SH$, Cs_2CO_3, DMF, 25 °C, 16 h.

The introduction of a carbamate linker at the 3'–position of nucleosides is possible starting from commercially available isocyanates such as phenyl isocyanates,[18] however this approach appears unsuitable for macromolecules containing unprotected polyhydroxyl and polyamino functionalities. Another commonly applied method to generate carbamates is the activation of the 3'-OH group with CDI,[19] but we felt that the strictly anhydrous conditions required by this protocol would not be compatible with the poor solubility of most aminoglycosides. A solution was found by choosing as

activating/leaving group 4-nitrophenyl carbonate,[20] which is electrophilic enough to react with amino functionalities, whilst still reasonably stable in aqueous media. Compound **5.20** was prepared in a series of straightforward steps, involving phosphorylation at the 5'-position of thymidine after selective protection−deprotection at the 3'- and 5'-position with a MMTr and benzoyl group respectively and final 3'-OH activation with 4-nitrophenyl chloroformate in a controlled way in the presence of pyridine/DCM. With both coupling units now available, we proceeded to react 3'-*O*-(4-nitrophenylcarbonate)−TMP **5.20** with *N*-Boc kanamycin A cysteamine **5.19** (Scheme 5-6A).



Scheme 5-6. Synthesis of 3'-carbamoyl kanamycin A−TMP conjugates **5.22** and **5.24**. Reagents and conditions: (a) Et_3N, 1,4-dioxane/water, 0 °C to rt, 18 h; (b) 20% $Pd(OH)_2/C$, H_2, EtOH/H_2O (8:2), rt, 24 h; (c) TFA, thioanisole, H_2O, rt, 5 h.

An initial attempt to synthesize **5.21** using previously optimized conditions in the presence of triethylamine as base and DCM/DMF (1:1) as solvent, afforded the desired conjugate in poor yield. Replacement of the solvent system with a mixture of 1,4-dioxane/water (3:1), surprisingly resulted in an improved yield (90%) (Table 3-1). Removal of the dibenzyl phosphate protection of **5.21** was

achieved by catalytic hydrogenolysis using an excess of 20% $Pd(OH)_2/C$ to avoid potential catalyst poisoning due to the sulfide moiety. Without further purification, the *N*-Boc protected intermediate **5.21** was converted to compound **5.22** upon treatment with aqueous TFA, in good yield and high purity after preparative RP-HPLC purification.

As the convergent coupling of complex fragments is important to rapidly generate a variety of conjugates, we decided to further explore the versatility of this reaction process. Since a single aminomethyl group is present in kanamycin A, the regioselective formation of a mono-carbamate linker seemed feasible without prior protection of the other amino groups. The 6'-amino group of kanamycin A is known to be the most reactive following the series 6'-NH_2 > 1-NH_2 > 3-NH_2 > 3''-NH_2).[21, 22] When the compound **5.20** was allowed to react directly with kanamycin A **5.16** utilising the general protocol for carbamate formation (Scheme 3-6B), intermediate **5.23** was obtained in good yield. Subsequent hydrogenolysis in the presence of 20% $Pd(OH)_2/C$ and preparative RP-HPLC purification yielded compound **5.24** in good yield and high purity.

Table 5-1. Reaction circumstances for the synthesis of compounds **5.21** and **5.23** from **5.20**.

Entry	Solvent	Temp	Time (h)	Yield of **21**[a] (%)	Yield of **23**[a] (%)
1	DCM : DMF (1:1)	rt	60	48	-
2	DMF	rt	48	55	42
3	Dioxane : H_2O (3:1)	0 °C to rt	16	90	82

[a] Isolated yield

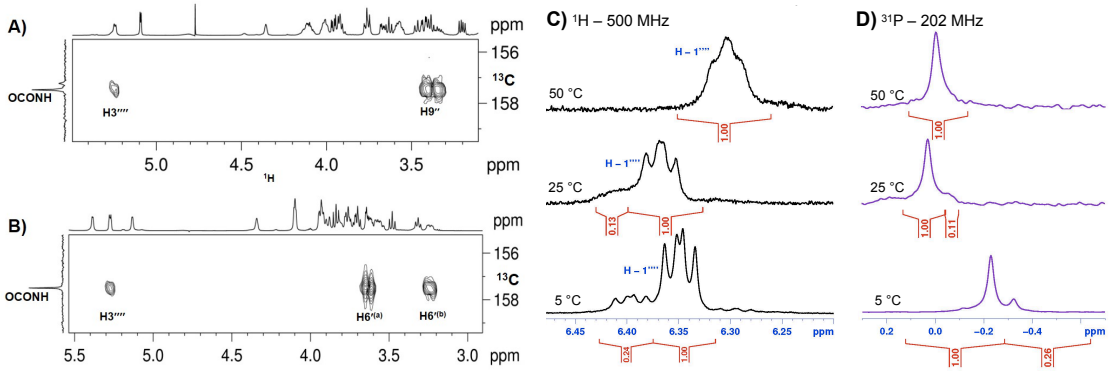


Figure 5-4. A) Key HMBC correlation for compound **5.22**; B) key HMBC correlation NMR spectrum of compound **5.24**; C) and D) temperature-dependent NMR (^1H and ^{31}P) study of compound **5.22**. All NMR spectra were recorded in D_2O.

Although the reactivity order of the various hydroxyl and amino groups of kanamycin A is well documented, we thought necessary to confirm the structures of the final analogues. The peak assignment for almost all proton and carbon signals of kanamycin A−TMP conjugates **5.22** and **5.24** was unambiguously determined (Table 5-2), as well as the alignment of the carbamoyl linkage between the 9'' and 6' positions (Figure 5-4A and B), by means of two-dimensional NMR

spectroscopy in D_2O solution at 25 °C. The presence of a distinctive signal at 157.5 ppm in the carbonyl region of the carbon NMR spectrum, proton integration ratio along with HRMS mass spectrometry confirmed the formation of a single carbamate linkage at the expected positions. While performing the ^{1}H and ^{31}P NMR spectroscopic analysis of compound **3.22** (Figure 5-4C and D), the appearance of a minor conformation was observed at low temperature (5 °C), which disappeared when the temperature was raised (50 °C), indicating the presence of a single conformer. This phenomenon may arise due to the formation of rotamers, as the inherent flexibility around glycosidic bonds of kanamycin A−TMP conjugates allows them to easily adopt an array of conformations. [23]

5.3. CONCLUSION

In summary, we have accomplished the synthesis of a variety of biotin−TMP and kanamycin A−TMP conjugates in a straightforward manner starting from previously optimized 3'-modified TMP intermediates and featuring in turn an oxymethyleneoxy ester, a carboxy ester, a carboxamide, a carbamoyl and a thioethyl carbamoyl group. Particularly, kanamycin A-dTMP conjugates connected through carbamate linker, were accomplished in a convergent manner considering the regioselectivity of the primary hydroxyl and amino functionality.

5.4. EXPERIMENTAL SECTION

General information

For all reactions, analytical grade solvents were used. Kanamycin A . H_2SO_4 was purchased from Sigma-Aldrich (containing kanamycin B as a 5% impurity). Kanamycin A. H_2SO_4 salt was transformed into its free amino form by treatment with Amberlite IRA-400 ion exchange resin. All moisture-sensitive reactions were carried out using oven-dried glassware (135 ºC) under a nitrogen or argon atmosphere. Reaction temperatures are reported as bath temperatures. Pre-coated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ = 254 nm). Products were purified by flash chromatography on ICN silica gel 63-200, 60 Å. All final compounds were purified by preparative RP-HPLC using a gradient of H_2O and MeCN, both contains either 50 mmol TEAB or 0.1% TFA as eluent buffer as mentioned. ^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on 300, 500 or 600 MHz Bruker spectrometers. Final compounds were characterized using 2D NMR (H-COSY, HSQC, HMBC, TOCSY, NOESY) techniques. The ^{1}H and ^{13}C chemical shifts were referenced to residual solvent signals relative to TMS as internal standard wherever applied. Coupling constants *J* [Hz] were directly taken from the spectra. Splitting patterns are designated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet) and br (broad). High-resolution mass spectra were acquired on a quadrupole orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 μL/min and spectra were obtained in positive (or negative) ionization mode with a resolution of 15000 (FWHM) using leucine enkephalin as lock mass.

5'-*O*-(*tert*-Butyl diphenyl silane)-3'-*O*-(biotinmethyloxyester)-thymidine (5.2)

A 1M solution of SO_2Cl_2 in CH_2Cl_2 (2.07 mL, 2.07 mmol) was added to a stirred solution of **5.1** (0.935 g, 0.107 mmol) in dry CH_2Cl_2 (20 mL) at 0 °C. The reaction mixture was allowed to warm slowly to room temperature over 2 h, then it was concentrated using a rotary evaporator (bath

temperature 10-15 °C) under reduced pressure to give a 3'-*O*-chloromethylthymidine derivative as a light yellowish foam, which was used without further purification.

In a separate round-bottomed flask, biotin (0.634 g, 2.593 mmol) was suspended in dry CH_2Cl_2 (20 mL), and DBU (0.38 mL, 2.51 mmol) was then added. After 15 min, the solution was added to the crude 3'-*O*-chloromethylthymidine derivative (redissolved in dry CH_2Cl_2, 20 mL), and the reaction mixture was kept stirring at room temperature for 24 h. The mixture was then diluted with CH_2Cl_2 (100 mL), and 1M aq. acetic acid (4 mL) was added with vigorous stirring. The aqueous layer was discarded, and the organic layer was washed with saturated aq. $NaHCO_3$ (2 x 50 mL) and brine, dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH, 99:1, v/v; 98:2, v/v; 97:3, v/v) to give **5.2** (0.91 g, 71%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 11.9 (s, 1H, N*H*-Thy), 7.68-7.65 (m, 4H, Ar-*H* Ph), 7.51 (s, 1H, H-6), 7.46-7.37 (m, 6H, Ar-*H* Ph), 6.94 (s, 1H, N*H* Bio), 6.91 (s, 1H, N*H* Bio), 6.27 (dd, *J* = 8.6, 5.2 Hz, 1H, H-1'), 5.71 (m, 1H, H-g Bio), 4.92 (m, 1H, H-f Bio), 4.59-4.54 (m, 1H, H-3'), 4.38-4.35 (m, 2H, H-5' and H-5''), 4.12-4.10 (m, 1H, H-4'), 4.03-3.82 (m, 2H, OCH_2O), 3.21-3.14 (m, 1H, H-e Bio), 2.96-2.82 (m, 2H, H-h Bio and H-h' Bio), 2.74-2.68 (m, 1H, H-2'), 2.49-2.28 (m, 2H, H-a Bio), 2.14-2.05 (m, 1H, H-2''), 1.83-1.65 (m, 4H, H-b and H-d Bio), 1.62 (s, 3H, CH_3), 1.55-1.40 (m, 2H, H-c Bio), 1.11 (s, 9H, ^{t}Bu); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.6 (OCO), 165.0 (C-4), 164.3 (NHCONH), 151.4 (C-2), 135.4 (Ar-C), 135.2 (Ar-C), 134.3 (C-6), 132.6 (Ar-C), 132.2 (Ar-C), 130.1 (Ar-C), 130.0 (Ar-C), 128.0 (Ar-C), 127.9 (Ar-C), 111.2 (C-5), 88.1 (OCH_2O), 85.1 (C-1'), 84.5 (C-4'), 81.6 (C-3'), 64.0 (C-5'), 62.1 (C-f Bio), 60.3 (C-g Bio), 55.4 (C-e Bio), 40.7 (C-h, h' Bio), 38.9 (C-2'), 33.8 (C-a Bio), 28.0 (C-c Bio), 27.9 (C-d Bio), 26.9 (^{t}Bu), 24.5 (C-b Bio), 19.2 (1C ^{t}Bu), 12.1 (CH_3-Thy); HRMS for $C_{37}H_{48}N_4O_8SSi$ $[M+H]^+$ calcd.: 737.3035, found: 737.3045.

3'-*O*-(biotinmethyloxyester)-thymidine (5.3)

Triethylamine trihydrofluoride (0.76 mL, 4.67 mmol) was added to a stirred solution of **5.2** (0.86 g, 1.17 mmol) in THF (20 mL) in a falcon tube at room temperature. The reaction mixture was stirred at room temp. for 72 h, and then it was concentrated under reduced pressure. The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH, 99:1, v/v; 96:4, v/v; 93:7, v/v) to give **5.3** (0.49 g, 84%) as a white solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.34 (s, 1H, N*H*-Thy), 7.68 (s, 1H, H-6), 6.42 (s, 1H, N*H* Bio), 6.37 (s, 1H, N*H* Bio), 6.01 (app t, *J* = 7.0 Hz, 1H, H-1'), 5.30 (dd, *J* = 20.2, 6.6 Hz, 2H, OCH_2O), 5.13 (t, *J* = 5.1 Hz, 1H, 5'-OH), 4.39-4.36 (m, 1H, H-3'), 4.33-4.28 (m, 1H, H-g Bio), 4.15-4.10 (m, 1H, H-f Bio), 3.93-3.90 (m, 1H, H-4'), 3.61-3.56 (m, 2H, H-5' and H-5''), 3.13-3.06 (m, 1H, H-e Bio), 2.82 (dd, *J* = 12.4, 5.1 Hz, 1H, H-h Bio), 2.57 (d, *J* = 12.4 Hz, 1H, H-h' Bio), 2.36 (t, *J* = 7.4 Hz, 1H, and H-a Bio), 2.27-2.18 (m, 2H, H-2' and H-2''), 1.77 (s, 3H, CH_3), 1.66-1.45 (m, 4H, H-b and H-d Bio), 1.42-1.29 (m, 2H, H-c Bio); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 172.4 (OCO), 163.6 (C-4), 162.7 (NHCONH), 150.5 (C-2), 135.9 (C-6), 109.5 (C-5), 87.4 (OCH_2O), 84.8 (C-1'), 83.6 (C-4'), 79.4 (C-3'), 61.1 (C-5'), 61.0 (C-f Bio), 59.2 (C-g Bio), 55.3 (C-e Bio), 40.2 (C-h, h' Bio, merged with DMSO), 36.9 (C-2'), 33.3 (C-a Bio), 28.0 (C-c Bio), 27.9 (C-d Bio), 24.2 (C-b Bio), 12.2 (CH_3-Thy); HRMS for $C_{21}H_{30}N_4O_8S$ $[M-H]^-$ calcd.: 497.1711, found: 497.1715.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(methylthiomethyl)-thymidine (5.6)

Acetic anhydride (10.9 mL) and acetic acid (3.5 mL) were added to a stirred solution of **5.5** (2.40 g, 4.78 mmol) in DMSO (15.5 mL). The reaction mixture was stirred at room temperature for 30 h, and then it was concentrated under reduced pressure. The residue was neutralized with saturated aq. $NaHCO_3$ (230 mL), and the mixture was extracted with ethyl acetate (3 x 170 mL). The combined organic extracts were washed with brine, dried over Na_2SO_4, filtered and concentrated under reduced

pressure. The crude residue was purified by column chromatography on silica gel (gradient hexane/EtOAc, 2:1, v/v; 1:1, v/v; 1:3, v/v) to give **5.6** (1.64 g, 61%) as a colorless foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.22 (br s, 1H, N*H*), 7.37-7.35 (m, 11H, H-6 and Ar-*H* OBn), 6.26 (dd, *J* = 8.2, 5.8 Hz, 1H, H-1'), 5.11-5.00 (m, 4H, 2 x OCH_2Ph), 4.64-4.51 (m, 2H, OCH_2S), 4.35-4.31 (m, 1H, H-3'), 4.17-4.09 (m, 3H, H-4', H-5' and H-5''), 2.34-2.26 (m, 1H, H-2'), 2.11 (s, 3H, SCH_3), 1.92-1.81 (m, 4H, H-2'' and CH_3-Thy); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 163.4 (C-4), 150.2 (C-2), 135.6 (d, $^3J_{C,P}$ = 6.1 Hz, 1C of OCH_2Ph), 135.3 (C-6), 129.0 (Ar-C), 128.9 (Ar-C), 128.2 (Ar-C), 111.5 (C-5), 85.0 (C-1'), 82.8 (d, $^3J_{C,P}$ = 8.2 Hz, C-4'), 76.3 (C-3'), 74.1 (OCH_2S), 69.9-69.8 (2 x d, $^2J_{C,P}$ = 3.5 Hz, 2 x OCH_2Ph), 67.0 (d, $^2J_{C,P}$ =5.7 Hz, C-5'), 37.5 (C-2'), 14.0 (SCH_3), 12.5 (CH_3-Thy); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.5; HRMS for $C_{26}H_{31}N_2O_8PS$ $[M-H]^-$ calcd.: 561.1466, found: 561.1461.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(biotinmethyloxyester)-thymidine (5.4)

Following a similar procedure as the one used for the synthesis of **5.2**, compound **5.4** was obtained (41.0 mg, 51%) as a colorless foam, starting from **5.6** (60.0 mg, 0.107 mmol), a 1M solution of SO_2Cl_2 in CH_2Cl_2 (0.13 mL, 0.128 mmol) in dry CH_2Cl_2 (4 mL) and biotin (39.1 mg, 0.160 mmol), DBU (0.023 mL, 0.155 mmol) in dry CH_2Cl_2 (15 mL). The crude residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH, 99:1, v/v; 97:3, v/v; 93:7, v/v). ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.76 (s, 1H, N*H*-Thy), 7.35-7.37 (m, 11H, H-6 and Ar-*H* OBn), 6.81 (br s, 1H, N*H* Bio), 6.31 (app t, *J* = 6.6 Hz, 1H, H-1'), 6.18 (br s, 1H, N*H* Bio), 5.13-4.99 (m, 6H, OCH_2O and 2 x OCH_2Ph), 4.53-4.49 (m, 1H, H-3'), 4.38-4.35 (m, 1H, H-g Bio), 4.26-4.12 (m, 3H, H-f Bio, H-5' and H-5''), 4.05-4.03 (m, 1H, H-4'), 3.16-3.12 (m, 1H, H-e Bio), 2.89 (dd, *J* = 12.2, 7.2 Hz, 1H, H-h Bio), 2.79 (d, *J* = 12.2 Hz, 1H, H-h' Bio), 2.58-2.52 (m, 1H, H-2'), 2.40-2.27 (m, 3H, H-2'' and H-a Bio), 1.80 (s, 3H, CH_3), 1.76-1.69 (m, 4H, H-b and H-d Bio), 1.52-1.43 (m, 2H, H-c Bio); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.7 (OCO), 164.7 (C-4), 164.2 (NHCONH), 150.7 (C-2), 135.5 (d, $^3J_{C,P}$ = 8.0 Hz, 1C of OCH_2Ph), 134.5 (C-6), 128.8 (Ar-C), 128.7 (Ar-C), 128.1 (Ar-C), 111.3 (C-5), 88.1 (OCH_2O), 84.8 (C-1'), 83.1 (d, $^3J_{C,P}$ = 8.1 Hz, C-4'), 81.3 (C-3'), 69.9-69.8 (2 x d, $^2J_{C,P}$ = 5.2 Hz, 2 x OCH_2Ph), 67.0 (d, $^2J_{C,P}$ = 5.9 Hz, C-5'), 62.3 (C-f Bio), 60.2 (C-g Bio), 55.6 (C-e Bio), 40.2 (C-h, h' Bio), 38.6 (C-2'), 33.9 (C-a Bio), 28.2 (C-c Bio), 28.0 (C-d Bio), 24.8 (C-b Bio), 12.5 (CH_3-Thy); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.4; HRMS for $C_{35}H_{43}N_4O_{11}PS$ $[M+Na]^+$ calcd.: 781.2279, found: 781.2288.

3'-*O*-(Biotinmethyloxyester)-thymidine monophosphate triethylammonium salt (5.7)

10% Pd/C (Degussa, 15.0 mg, 50% w/w) was added to a degassed stirred solution of **5.4** (30.0 mg, 0.039 mmol) and Et_3N (0.011 mL, 0.079 mmol) in MeOH (10 mL), and the mixture was subjected to hydrogenation at atmospheric pressure using a balloon filled with H_2 for 24 h. The catalyst was removed by filtration through a cellulose filter (0.45 μm) and the filtrate was concentrated under reduced pressure (bath temp 10-15 °C). The resulting crude residue was purified by RP-HPLC [TEAB (triethylammonium bicarbonate; 50 mmol) in H_2O/MeCN, 98:2; and TEAB (50 mmol) in H_2O/MeCN, 50:50]. The collected eluate was freeze-dried repeatedly until constant mass to obtain the triethylammonium salt of **5.7** (22.2 mg, 72%) as a white solid. ^{1}H NMR (500 MHz, D_2O) δ = 7.80 (s, 1H, H-6), 6.28 (app t, *J* = 7.3 Hz, 1H, H-1'), 5.46-5.32 (m, 2H, OCH_2O), 4.63-4.62 (m, 1H, H-3'), 4.56-4.54 (m, 1H, H-g Bio), 4.38-4.36 (m, 1H, H-f Bio), 4.25 (br s, 1H, H-4'), 3.94-3.88 (m, 2H, H-5' and H-5''), 3.29-3.25 (m, 1H, H-e Bio), 2.92 (dd, *J* = 13.0, 5.0 Hz, 1H, H-h), 2.70 (d, 1H, *J* = 13.0 Hz, H-h' Bio), 2.45 (t, 2H, *J* = 7.2 Hz, H-a Bio), 2.42-2.40 (m, 1H, H-2' and H-2''), 1.88 (s, 3H, CH_3), 1.73-1.52 (m, 4H, H-b and H-d Bio), 1.43-1.36 (m, 2H, H-c Bio) ; ^{13}C NMR (125 MHz, D_2O) δ = 175.5 (OCO), 168.3 (C-4), 164.8 (NHCONH), 152.9 (C-2), 137.0 (C-6), 111.4 (C-5), 88.1 (OCH_2O), 84.3 (C-1'), 83.8 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 80.7 (C-3'), 63.3 (d, $^2J_{C,P}$ = 3.9 Hz, C-5'), 61.5 (C-f Bio), 59.7 (C-g Bio), 54.8 (C-e Bio), 39.2 (C-h, h' Bio), 36.6 (C-2'), 33.1 (C-a Bio), 27.4 (C-c),

27.1 (C-d Bio), 23.4 (C-b Bio), 11.5 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 3.4; HRMS for $C_{21}H_{31}N_4O_{11}PS$ $[M-H]^-$ calcd.: 577.1375, found: 577.1375.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(biotincarboxamide)-thymidine (5.9)

DCC (127.4 mg, 0.618 mmol) was added to a stirred solution of **5.8** (200.0 mg, 0.386 mmol), biotin (108.6 mg, 0.444 mmol), and DMAP (2.0 mg, 0.016 mmol) in a mixture of dry DMF (2.0 mL) and dry CH_2Cl_2 (10 mL) at 0 °C. The reaction mixture was allowed to warm slowly to room temperature and stirred for 24 h. All the volatiles were removed under reduced pressure and the resulting residue was purified by column chromatography on silica gel (gradient CH_2Cl_2/MeOH, 99:2, v/v; 97:4, v/v; 94:7, v/v) to give **5.9** (201.0 mg, 70%) as a colourless foam. ^{1}H NMR (300 MHz, $CDCl_3$ + CD_3OD) δ = 7.41 (d, *J* = 1.0 Hz, 1H, H-6), 7.39-7.34 (m, 10H, Ar-*H* OBn), 6.27 (dd, *J* = 8.9, 5.6 Hz, 1H, H-1'), 5.12-5.00 (m, 4H, 2 x OC*H*$_2$Ph), 4.55-4.48 (m, 2H, H-3' and H-g Bio), 4.33-4.29 (m, 2H, H-4' and H-f Bio,), 4.23-4.22 (m, 2H, H-5' and H-5''), 3.21-3.15 (m, 1H, H-e Bio), 2.91 (dd, *J* = 12.8, 4.9 Hz, 1H, H-h Bio), 2.73 (d, *J* = 12.8 Hz, 1H, H-h' Bio), 2.54-2.48 (m, 1H, H-2'), 2.14 (t, *J* = 6.7 Hz, 3H, H-a Bio), 1.94-1.84 (m, 1H, H-2''), 1.81 (d, *J* = 0.9 Hz, 3H, C*H*$_3$), 1.75-1.61 (m, 4H, *H*-b and *H*-d Bio), 1.51-1.43 (m, 2H, H-c Bio) ; ^{13}C NMR (75 MHz, $CDCl_3$ + CD_3OD) δ = 172.8 (ONHCO), 165.1 (C-4), 164.9 (NHCONH), 151.3 (C-2), 136.0 (C-6), 135.6 (d, $^3J_{C,P}$ = 5.7 Hz, 1C of OCH_2Ph), 129.4 (Ar-C), 129.1 (Ar-C), 128.5 (Ar-C), 111.8 (C-5), 85.4 (C-3'), 85.3 (C-1'), 81.6 (d, $^3J_{C,P}$ = 8.3 Hz, C-4'), 70.5-70.4 (d, $^2J_{C,P}$ = 5.6 Hz, 2 x OCH_2Ph), 68.1 (d, $^2J_{C,P}$ = 5.8 Hz, C-5'), 62.4 (C-f Bio), 60.6 (C-g Bio), 55.9 (C-e Bio), 40.6 (C-h, h' Bio), 36.4 (C-2'), 32.8 (C-a Bio), 28.7 (C-c Bio), 28.5 (C-d Bio), 25.5 (C-b Bio), 12.4 (CH_3-Thy); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -1.0; HRMS for $C_{34}H_{42}N_5O_{10}PS$ $[M+Na]^+$ calcd.: 766.2282, found: 766.2302.

3'-*O*-(Biotincarboxamide)-thymidine monophosphate triethylammonium salt (5.10)

Following a similar procedure as the one used for the synthesis of **5.7**, the triethylammonium salt of compound **5.10** was obtained as a white solid (169.0 mg, 82%), starting from **5.9** (200.0 mg, 0.269 mmol), Et_3N (0.075 mL, 0.538 mmol) and 10% Pd/C Degussa (100 mg, 50% w/w) in MeOH (20 mL). ^{1}H NMR (500 MHz, D_2O) δ = 7.90 (s, 1H, H-6), 6.40 (dd, *J* = 9.1, 5.7 Hz, 1H, H-1'), 4.73-4.72 (m, 1H, H-3'), 4.56-4.54 (m, 1H, H-g Bio), 4.38-4.35 (m, 1H, H-f Bio), 4.32 (br s, 1H, H-4'), 3.92-3.91 (m, 2H, H-5' and H-5''), 3.32-3.28 (m, 1H, H-e Bio), 2.93 (dd, *J* = 13.0, 4.9 Hz, 1H, H-h), 2.69 (d, 1H, *J* = 13.0 Hz, H-h' Bio), 2.48-2.44 (m, 1H, H-2'), 2.40-2.34 (m, 1H, H-2''), 2.15 (t, 2H, *J* = 7.0 Hz, H-a Bio), 1.89 (s, 3H, C*H*$_3$), 1.72-1.67 (m, 1H, H-b Bio), 1.65-1.60 (m, 2H, H-d Bio), 1.59-1.51 (m, 1H, H-b' Bio), 1.40-1.34 (m, 2H, H-c Bio) ; ^{13}C NMR (125 MHz, D_2O) δ = 172.8 (ONHCO), 166.9 (C-4), 164.8 (NHCONH), 151.8 (C-2), 137.3 (C-6), 111.4 (C-5), 84.9 (C-3'), 84.3 (C-1'), 82.1 (d, $^3J_{C,P}$ = 8.9 Hz, C-4'), 63.6 (d, $^2J_{C,P}$ = 3.9 Hz, C-5'), 61.4 (C-f Bio), 59.7 (C-g Bio), 54.7 (C-e Bio), 39.2 (C-h, h' Bio), 34.8 (C-2'), 31.8 (C-a Bio), 27.3 (C-c Bio), 27.0 (C-d Bio), 24.5 (C-b Bio), 11.3 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 3.4; HRMS for $C_{20}H_{30}N_5O_{10}PS$ $[M-H]^-$ calcd.: 562.1378, found: 562.1382.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(biotinmethyloxycarboxamide)-thymidine (5.12)

Following a similar procedure as the one used for the synthesis of **5.9**, compound **5.12** was obtained as a colorless foam (214.0 mg, 76%), starting from **5.11** (200 mg, 0.365 mmol), biotin (102.6 mg, 0.420 mmol), DCC (120.6 mg, 0.584 mmol) and DMAP (1.8 mg, 0.0146 mmol) in a mixture of dry DMF (2.0 ml) and dry CH_2Cl_2 (10 ml). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 11.34 (br s, 1H, N*H*), 11.00 (br s, 1H, N*H*), 7.50 (s, 1H, H-6), 7.35-7.35 (m, 10H, Ar-*H* OBn), 6.41 (br s, 1H, N*H* Bio), 6.35 (br s, 1H, N*H* Bio), 6.15 (dd, *J* = 8.2, 5.9 Hz, 1H, H-1'), 5.06 (d, *J* = 8.0 Hz, 4H, 2 x OC*H*$_2$Ph), 4.87-4.81 (m, 2H, OC*H*$_2$O), 4.51-4.50 (m, 1H, H-3'), 4.33-4.29 (m, 3H, H-g Bio, H-f Bio and H-4'), 4.19-4.10 (m, 2H, H-5' and H-5''), 3.10-3.04 (m, 1H, H-e Bio), 2.81 (dd, *J* = 12.6, 5.2 Hz, 1H, H-h Bio),

2.57 (d, J = 12.6 Hz, 1H, H-h' Bio), 2.35-2.28 (m, 1H, H-2'), 2.17-2.08 (m, 1H, H-2''), 2.00 (t, J = 7.0 Hz, 3H, H-a Bio), 1.69 (s, 3H, CH_3), 1.63-1.42 (m, 4H, H-b and H-d Bio), 1.36-1.27 (m, 2H, H-c Bio) ; ^{13}C NMR (75 MHz, DMSO-d_6) δ = 169.6 (ONHCO), 163.6 (C-4), 162.7 (NHCONH), 150.4 (C-2), 135.9 (d, $^3J_{C,P}$ = 6.9 Hz, 1C of OCH_2Ph), 135.6 (C-6), 128.5 (Ar-C), 128.4 (Ar-C), 127.8 (Ar-C), 109.9 (C-5), 97.0 (OCH_2O), 84.0 (C-1'), 82.3 (d, $^3J_{C,P}$ = 7.6 Hz, C-4'), 77.1 (C-3'), 68.7 (d, $^2J_{C,P}$ = 5.4 Hz, 2 x OCH_2Ph), 67.1 (d, $^2J_{C,P}$ = 5.0 Hz, C-5'), 61.0 (C-f Bio), 59.2 (C-g Bio), 55.3 (C-e Bio), 39.5 (C-h, h' Bio), 36.2 (C-2'), 32.0 (C-a Bio), 28.1 (C-c Bio), 28.0 (C-d Bio), 24.9 (C-b Bio), 12.0 (CH_3-Thy); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.9; HRMS for $C_{35}H_{44}N_5O_{11}PS$ $[M+Na]^+$ calcd.: 796.2388, found: 796.2403.

3'-*O*-(Biotinmethyloxycarboxamide)-thymidine monophosphate triethylammonium salt (5.13)

Following a similar procedure as the one used for the synthesis of **5.7**, the triethylammonium salt of compound **5.13** was obtained as a white solid (170.6 mg, 79%), starting from **5.12** (210 mg, 0.271 mmol), Et_3N (0.075 mL, 0.543 mmol) and 10% Pd/C Degussa (105 mg, 50% w/w) in MeOH (20 mL). ^{1}H NMR (500 MHz, D_2O) δ = 7.85 (d, J = 1.1 Hz, 1H, H-6), 6.31 (app t, J = 7.3 Hz, 1H, H-1'), 4.98 (s, 2H, OCH_2O), 4.71-4.70 (m, 1H, H-3'), 4.57-4.54 (m, 1H, H-g Bio), 4.39-4.37 (m, 1H, H-f Bio), 4.28-4.27 (m, 1H, H-4'), 3.97-3.90 (m, 2H, H-5' and H-5''), 3.31-3.27 (m, 1H, H-e Bio), 2.94 (dd, J = 13.0, 5.0 Hz, 1H, H-h), 2.71 (d, 1H, J = 13.0 Hz, H-h' Bio), 2.42-2.39 (m, 1H, H-2' and H-2''), 2.20 (t, 2H, J = 7.0 Hz, H-a Bio), 1.89 (d, J = 1.1 Hz, 3H, CH_3), 1.70-1.54 (m, 1H, H-b and H-d Bio), 1.39-1.37 (m, 2H, H-c Bio) ; ^{13}C NMR (125 MHz, D_2O) δ = 172.9 (ONHCO), 166.2 (C-4), 164.8 (NHCONH), 151.3 (C-2), 137.2 (C-6), 111.4 (C-5), 96.8 (OCH_2O), 84.3 (C-1'), 83.8 (d, $^3J_{C,P}$ = 8.7 Hz, C-4'), 78.4 (C-3'), 63.4 (d, $^2J_{C,P}$ = 3.7 Hz, C-5'), 61.5 (C-f Bio), 59.7 (C-g Bio), 54.8 (C-e Bio), 39.2 (C-h, h' Bio), 36.1 (C-2'), 31.5 (C-a Bio), 27.2 (C-c Bio), 27.1 (C-d Bio), 24.2 (C-b Bio), 11.2 (CH_3-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 2.8; HRMS for $C_{21}H_{32}N_5O_{11}PS$ $[M-H]^-$ calcd.: 592.1484, found: 592.1480.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(biotinester)-thymidine (5.14)

Following a similar procedure as the one used for the synthesis of **5.9**, compound **5.14** was obtained (370.0 mg, 74%) as a colorless foam, starting from **5.5** (345 mg, 0.686 mmol), biotin (201.0 mg, 0.823 mmol), DCC (198.0 mg, 0.960 mmol) and DMAP (8.0 mg, 0.068 mmol) in a mixture of dry DMF (3 ml) and dry CH_2Cl_2 (15 ml). ^{1}H NMR (300 MHz, $CDCl_3$) δ = 10.95 (br s, 1H, NH Thy), 7.40 (s, 1H, H-6), 7.34-7.33 (m, 10H, Ar-H OCH_2Ph), 6.76 (br s, 1H, NH Bio), 6.50 (br s, 1H, NH Bio), 6.26 (app t, J = 6.7 Hz, 1H, H-1'), 5.13-5.00 (m, 5H, H-3' and OCH_2Ph), 4.51-4.50 (m, 1H, H-g Bio), 4.30-4.22 (m, 1H, H-f Bio, H-5' and H-5''), 4.11 (br s, 1H, H-4'), 3.17-3.15 (m, 1H, H-e Bio), 2.91-2.88 (m, 1H, H-h Bio), 2.78-2.74 (m, 1H, H-h' Bio), 2.35-2.25 (m, 3H, H-a Bio and H-2'), 1.97-1.94 (m, 1H, H-2''), 1.81 (s, 3H, CH_3), 1.69-1.60 (m, 4H, H-b and H-d Bio), 1.46-1.44 (m, 2H, H-c Bio) ; ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.7 (OCO), 164.3 (C-4), 164.2 (NHCONH), 150.5 (C-2), 135.1 (d, $^3J_{C,P}$ = 6.1 Hz, 1C of OCH_2Ph), 134.7 (C-6), 128.5 (Ar-C), 128.3 (Ar-C), 127.7 (Ar-C), 111.2 (C-5), 84.1 (C-1'), 82.4 (d, $^3J_{C,P}$ = 8.1 Hz, C-4'), 74.1 (C-3'), 69.4 (app t, $^2J_{C,P}$ = 4.7 Hz, 2 x OCH_2Ph), 66.8 (d, $^2J_{C,P}$ = 4.7 Hz, C-5'), 61.7 (C-f Bio), 60.0 (C-g Bio), 55.2 (C-e Bio), 40.2 (C-h, h' Bio), 36.8 (C-2'), 33.4 (C-a Bio), 28.0 (C-c Bio), 27.8 (C-d Bio), 24.3 (C-b Bio), 12.0 (CH_3-Thy); ^{31}P NMR (121 MHz, $CDCl_3$) δ = -0.7; HRMS for $C_{34}H_{41}N_4O_{10}PS$ $[M+H]^+$ calcd.: 729.2353, found: 729.2354.

3'-*O*-(Biotinester)-thymidine monophosphate triethylammonium salt (5.15)

Following a similar procedure as the one used for the synthesis of **5.7**, the triethylammonium salt of compound **5.15** was obtained as a white solid (335.0 mg, 88%), starting from **5.14** (370.0 mg, 0.686 mmol), Et_3N (0.141 mL, 1.015 mmol) and 10% Pd/C (Degussa, 185.0 mg, 50% w/w) in MeOH (30 ml). ^{1}H NMR (600 MHz, DMSO-d_6) δ = 11.34 (br s, 1H, NH Thy), 7.91 (s, 1H, H-6), 6.51 (br s, 1H,

N*H* Bio), 6.40 (br s, 1H, N*H* Bio), 6.23 (dd, *J* = 9.0, 5.8 Hz, 1H, H-1'), 5.28-5.27 (m, 1H, H-3'), 4.32-4.30 (m, 1H, H-g Bio), 4.15-4.12 (m, 1H, H-f Bio), 4.07 (br s, 1H, H-4'), 3.90-3.83 (m, 2H, H-5' and H-5''), 3.13-3.10 (m, 1H, H-e Bio), 2.83-2.82 (m, 1H, H-h Bio merged with triethylamine), 2.58 (d, 1H, *J* = 12.4 Hz, H-h' Bio), 2.36 (t, 2H, *J* = 7.1 Hz, H-a Bio), 2.33-2.29 (m, 1H, H-2'), 2.21-2.17 (m, 1H, H-2''), 1.81 (s, 3H, CH_3), 1.66-1.62 (m, 1H, H-b Bio), 1.61-1.55 (m, 2H, H-d Bio), 1.51-1.44 (m, 1H, H-b' Bio), 1.39-1.32 (m, 2H, H-c Bio); ^{13}C NMR (150 MHz, DMSO-d_6) δ = 172.6 (OCO), 163.9 (C-4), 162.9 (NHCONH), 150.7 (C-2), 136.2 (C-6), 110.3 (C-5), 83.7 (C-1'), 83.5 (d, $^{3}J_{C,P}$ = 7.4 Hz, C-4'), 75.5 (C-3'), 64.2 (d, $^{2}J_{C,P}$ = 6.0 Hz, C-5'), 61.1 (C-f Bio), 59.3 (C-g Bio), 55.4 (C-e Bio), 39.6 (C-h,h' Bio merged with DMSO), 36.5 (C-2'), 33.4 (C-a Bio), 28.0 (C-c and C-d Bio), 24.5 (C-b Bio), 12.2 (CH_3-Thy); ^{31}P NMR (202 MHz, DMSO-d_6) δ = -0.1; HRMS for $C_{20}H_{29}N_4O_{10}PS$ $[M-H]^-$ calcd.: 547.1269, found: 547.1279.

***N*-Boc kanamycin A (5.17)**

To a stirred solution of kanamycin A **5.16** (1.8 g, 3.71 mmol) in a mixture of DMF:H_2O (4:1, 48 mL), was added di-*tert*-butyldicarbonate (5.12 mL, 22.29 mmol) and the solution was heated at 60 °C for 5h and then cooled to room temperature. After removal of all the volatiles, the resulting residue was washed with hexane (2 x 100 ml) and then with water (3 x 100 mL), and finally dried *in vacuo* to afford **5.17** (2.99 g, 91%) as a white solid. R_f = 0.4 (10% MeOH in DCM). ^{1}H NMR (300 MHz, DMSO-d_6) δ = 6.89 (br s, 1H), 6.58 (br s, 1H), 6.50 (d, *J* = 8.6 Hz, 1H), 6.34 (br s, 1H), 5.35 (d, *J* = 3.8 Hz, 1H), 5.26 (s, 1H), 4.89 (m, 4H), 4.68 (d, *J* = 6.2 Hz, 1H), 4.19 (t, *J* = 5.4 Hz, 1H), 3.82 (m, 1H), 3.61-3.35 (m, 9H), 3.30-3.17 (m, 6H), 3.07 (m, 1H), 1.80 (m, 1H), 1.40-1.35 (m, 37H); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 156.3, 156.1, 155.3, 154.9, 101.1, 97.8, 84.0, 80.3, 77.8, 77.7, 77.2, 75.0, 72.9, 72.7, 72.1, 70.5, 70.3, 70.1, 67.4, 60.3, 55.9, 50.0, 49.1, 41.4, 34.7, 28.3, 28.2, 28.1; HRMS [ESI$^+$] for $C_{38}H_{68}N_4O_{19}$ $[M+H]^+$calcd.: 885.4550, found: 885.4552.

***O*-TIBS-*N*-Boc kanamycin A (5.18)**

To a stirred solution of **5.17** (2.48 g, 2.81 mmol) in dry pyridine (50 mL) was added 2,4,6-triisopropylbenzenesulfonyl chloride (5.96 g, 19.67 mmol) and the solution was stirred at 25 °C for 16h. The reaction mixture was neutralized by adding 1N HCl and diluted with water. The aqueous layer was extracted with ethyl acetate (3 x 400 mL). The collected organic fractions were washed with brine, dried over Na_2SO_4 and concentrated *in vacuo*. The crude residue was purified by column chromatography over silica gel (3% MeOH in DCM) to afford **5.18** (1.91 g, 59%) as a white solid. R_f = 0.48 (10% MeOH in DCM). ^{1}H NMR (300 MHz, CD_3OD) δ = 7.28 (s, 1H), 6.58 (br s, 1H), 5.35 (d, *J* = 3.8 Hz, 1H), 5.04 (m, 2H), 4.38 (m, 2H), 4.14 (m, 3H), 3.71-3.31 (m, 12H), 3.18 (t, *J* = 9.4 Hz, 1H), 2.95 (m, 1H), 2.02 (m, 1H), 1.45-1.41 (m, 37H), 1.28-1.25 (m, 18H); ^{13}C NMR (75 MHz, CD_3OD) δ = 159.4, 159.2, 157.9, 157.7, 155.3, 152.3, 130.7, 124.9, 102.8, 99.8, 85.8, 80.9, 80.6, 80.4, 80.2, 76.8, 74.5, 73.9, 72.3, 71.9, 71.7, 71.6, 69.2, 57.4, 52.2, 50.9, 41.9, 36.0, 35.5, 30.8, 28.9, 28.8, 28.7, 25.2, 25.1, 23.9; HRMS [ESI$^+$] for $C_{53}H_{90}N_4O_{21}S$ $[M+H]^+$calcd.: 1151.5890, found: 1151.5895.

***N*-Boc kanamycin A cysteamine (5.19)**

To a stirred solution of **5.18** (0.9 g, 0.782 mmol) in DMF (30 mL) was added 2-mercaptoethylamine hydrochloride (0.67 g, 5.86 mmol) followed by cesium carbonate (2.55 g, 7.82 mmol) and the solution was stirred at 25 °C for 16h. DMF was partly removed *in vacuo* and then partitioned between water (150 ml) and ethyl acetate (300 ml). The aqueous layer was extracted with ethyl acetate (2 x 150 mL). The collected organic fractions were washed with brine, dried over Na_2SO_4 and concentrated *in vacuo*. The crude residue was purified by column chromatography over silica gel (5% MeOH in DCM) to afford **5.19** (0.45 g, 61%) as an off-white solid. R_f= 0.38 (10% MeOH in DCM). ^{1}H NMR (300 MHz, CD_3OD) δ = 5.13 (d, *J* = 3.5 Hz, 1H), 5.08 (d, *J* = 3.3 Hz, 1H), 4.17 (t, *J* = 6.7 Hz, 1H),

3.72-3.52 (m, 6H), 3.47-3.38 (m, 7H), 3.21 (m, 3H), 3.05 (m, 1H), 2.90-2.81 (m, 3H), 2.68 (dd, J = 14.5, 6.8 Hz, 1H), 2.12 (m, 1H), 1.46-1.43 (m, 37H); ^{13}C NMR (75 MHz, CD_3OD) δ = 159.4, 159.2, 158.1, 157.7, 102.6, 100.6, 85.2, 83.2, 80.6, 80.4, 80.2, 71.2, 74.5, 74.0, 73.8, 72.4, 72.1, 72.0, 71.8, 57.0, 52.0, 50.8, 41.8, 40.0, 35.9, 34.7, 31.4, 30.4, 28.9, 28.8, 28.8, 28.7, 24.2; HRMS [ESI$^+$] for $C_{40}H_{73}N_5O_{18}S$ [M+H]$^+$calcd.: 944.4744, found: 944.4733.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(*N*-Boc-kanamycin A cysteamine carbamate)-thymidine (5.21)

To a stirred solution of compound **5.19** (353 mg, 0.374 mmol) in 1,4-dioxane/water (3:1, 8 mL) was added **5.20** (250 mg, 0.374 mmol) followed by triethylamine (0.052 mL, 0.374 mmol) at 0 °C and then the solution was stirred at room temperature for 16 h. After completion of the reaction, the mixture was evaporated to dryness *in vacuo* and the residue was purified by column chromatography (4% MeOH in DCM) to obtain compound **5.21** (495 mg, 90%) as a white solid. R_f = 0.45 (10% MeOH in DCM). ^{1}H NMR (500 MHz, CD_3OD) δ = 7.47 (s, 1H), 7.36 (s, 10H), 6.25 (dd, J = 8.6, 5.8 Hz, 1H), 5.14-5.07 (m, 6H), 5.05 (br s, 1H), 4.29 (m, 2H), 4.20 (m, 2H), 3.71-3.60 (m, 4H), 3.55-3.48 (m, 4H), 3.42-3.32 (m, 6H), 3.19 (t, J = 9.3 Hz, 1H), 3.02 (dd, J = 13.9, 2.1 Hz, 1H), 2.75-2.62 (m, 3H), 2.33 (dd, J = 14.2, 5.5 Hz, 1H), 2.07 (m, 1H), 1.76 (s, 3H), 1.44-1.42 (m, 37H); ^{13}C NMR (125 MHz, CD_3OD) δ = 166.18, 159.5, 159.2, 158.0, 157.7, 152.3, 137.2, 137.1 (d, $^3J_{C,P}$ = 2.3 Hz), 137.0 (d, $^3J_{C,P}$ = 2.3 Hz), 130.0, 129.8, 129.3, 112.2, 102.7, 100.2, 86.3, 84.4 (d, $^3J_{C,P}$ = 7.8 Hz), 80.6, 80.5, 80.4, 80.2, 77.2, 76.0, 74.7, 74.1, 73.7, 72.5, 72.4, 72.3, 71.8, 71.2 (2 x d, $^2J_{C,P}$ = 5.8 Hz), 68.8 (d, $^2J_{C,P}$ = 6.1 Hz), 57.2, 50.9, 41.7, 38.2, 35.0, 33.9, 28.9, 28.9, 28.9, 28.8, 28.8, 12.6; ^{31}P NMR (202 MHz, CD_3OD) δ = -0.03; HRMS [ESI$^+$] for $C_{65}H_{98}N_7O_{27}PS$ [M+Na]$^+$calcd.: 1494.5861, found: 1494.5873.

3'-*O*-(kanamycin A cysteamine carbamate)-TMP (5.22)

To a stirred solution of **5.21** (200 mg, 0.136 mmol, 1 eq.) in EtOH/H_2O (8:2, 15 mL), was added 20% $Pd(OH)_2$/C (160 mg, 0.8 eq w/w) and evacuation was then carried out with hydrogen atmosphere replacements (3x). The reaction mixture was stirred at room temperature for 24h under an atmospheric pressure of hydrogen. After completion of the reaction, the catalyst was removed by filtration through a cellulose filter (0.45 μm) and the filtrate was concentrated under reduced pressure to obtain crude 3'-*O*-(*N*-Boc-kanamycin A cysteamine carbamate)-TMP (~175 mg, quant.). The so obtained residue was treated without further purification with 90% aq. TFA (5 mL) at 0 °C and then the solution was stirred at room temperature for 5h. After completion of the reaction, the mixture was evaporated to dryness *in vacuo*. The residue was coevaporated three times with toluene to remove residual TFA and to give crude 3'-*O*-(kanamycin A cysteamine carbamate)-TMP. The crude residue was purified by preparative RP-HPLC (0.1% TFA in 98% H_2O + 2% ACN and 0.1% TFA in 98% ACN + 2% H_2O). The collected eluates were freeze-dried repeatedly until constant mass to afford compound **5.22** (86 mg, 71% over two steps) as trifluoroacetic salt. ^{31}P NMR (202 MHz, D_2O) δ = -0.02; HRMS [ESI$^-$] for $C_{31}H_{54}N_7O_{19}PS$ [M-H]$^-$ calcd.: 890.2860, found: 890.2852.

5'-*O*-(Dibenzylphosphate)-3'-*O*-(kanamycin A carbamate)-thymidine (5.23)

Compound **5.23** was obtained according to the procedure for the preparation of compound **5.21** starting from kanamycin A **5.16** (100 mg, 0.206 mmol), compound **5.20** (138 mg, 0.206 mmol) and triethylamine (29 μL, 0.206 mmol). The crude residue was purified by column chromatography (IPA:H_2O:Et_3N 10:2:1) to give **5.23** as a white solid (171 mg, 82%). R_f = 0.45 (IPA:H_2O:Et_3N 6:2:2). ^{1}H NMR (300 MHz, D_2O) δ = 7.16 (s, 1H), 7.09 (s, 10H), 6.1 (app t, J = 6.0 Hz, 1H), 5.21 (d, J = 5.5 Hz, 1H), 5.01 (m, 2H), 4.84 (m, 4H), 4.19-3.89 (m, 4H), 3.81-3.08 (m, 14H), 2.99 (t, J = 10.4 Hz, 1H), 2.82 (t, J = 11.3 Hz, 1H), 2.31 (m, 1H), 2.04-1.80 (m, 2H), 1.52 (s, 3H), 1.33-1.19 (m, 1H); ^{13}C NMR (75 MHz, D_2O) δ = 165.3, 156.8, 151.0, 135.7, 134.8 (d, $^3J_{C,P}$ = 5.3 Hz), 128.5, 128.3, 127.7, 127.6, 127.1, 110.8, 99.9, 99.7, 86.8, 86.1, 84.6, 82.2, 74.2, 73.7, 72.4, 72.0, 71.5, 71.2, 70.6, 70.4,

69.6, 67.8, 67.0, 59.8, 54.2, 50.0, 48.9, 41.0, 36.2, 33.5, 11.5; ^{31}P NMR (121 MHz, D_2O) δ = -0.89; HRMS [ESI$^+$] for $C_{43}H_{61}N_6O_{20}P$ [M+H]$^+$calcd.: 1013.3751, found: 1013.3745.

3'-*O*-(kanamycin A carbamate)-TMP (5.24)

Compound **5.24** was obtained as a white solid (86.7 mg, 88%) starting from compound **5.23** (120 mg, 0.118 mmol) following the same hydrogenation procedure described for compound **5.22,** using 20% $Pd(OH)_2/C$ (24 mg, 0.2 eq w/w). ^{31}P NMR (202 MHz, D_2O) δ = -0.02; HRMS [ESI$^-$] for $C_{29}H_{49}N_6O_{20}P$ [M-H]$^-$ calcd.: 831.2666, found: 831.2666.

Table 5-2. ^{1}H and ^{13}C chemical shifts and peak assignments for compounds **5.22** and **5.24**.

	5.22* (pH = 2.20)$^\sharp$		**5.24*** (pH = 2.66)$^\sharp$	
No	δ_H (*J* in Hz)	δ_C	δ_H (*J* in Hz)	δ_C
1	3.57, m	49.7	3.58, m	49.4
2	2.52^{eq}, m 1.93^{ax}, m	27.5	2.54^{eq}, m 1.91^{ax}, m	27.5
3	3.58, m	47.6	3.55, m	48.0
4	3.91, m	78.6	3.84, app t (10)	79.4
5	3.92, m	72.7	3.93, m	72.4
6	3.75, app t (10)	83.9	3.75, m	83.5
1'	5.56, d (3.7)	96.6	5.38, d (3.2)	97.2
2'	3.67, dd (10, 3.7)	70.7	3.64, m	71.0
3'	3.76, app t (9.4)	72.1	3.71, m	72.4
4'	3.39, app t (9.7)	70.6	3.31, app t (9.4)	70.6
5'	4.01, m	68.7	3.81, m	71.6
6'	$3.43^{(a)}$, dd (14, 3.3) $3.20^{(b)}$, dd (14, 7.6)	40.2	$3.63^{(a)}$, m $3.24^{(b)}$, dd (14, 5.7)	41.1
1"	5.09, d (3.6)	100.5	5.13, d (2.6)	100.1
2"	3.96, dd (10.8, 3.5)	68.1	3.93, app t (10)	68.0
3"	3.46, app t (10.6)	54.7	3.69, app t (10)	65.4
4"	3.63, app t (10)	68.3	3.48, app t (10.6)	54.9
5"	4.01, m	72.0	3.92, m	73.0
6"	$3.08^{(a)}$, m $2.77^{(b)}$, m	32.6	$3.88^{(a)}$, m $3.76^{(b)}$, m	59.9
8"	2.77, m	32.2	-	-
9"	$3.34^{(a)}$, m $3.42^{(b)}$, m	39.9	-	-

OCONH	-	157.5	-	157.5
2'''	-	151.8	-	151.7
4'''	-	166.4	-	166.4
6'''	7.74, s	137.0	7.74, s	137.1
5-CH_3	1.91, s	11.5	1.91, s	11.5
1''''	6.35, dd (8.5, 6)	85.0	6.33, app t (7)	85.0
2'''' (a+b)	2.44, m	36.2	2.44, m	36.5
3''''	5.24, m	75.5	5.27, m	75.4
4''''	4.36, br s	83.7, d ($^3J_{C,P}$ = 8 Hz)	4.34, br s	83.4, d ($^3J_{C,P}$ = 8 Hz)
5'''' (a+b)	4.11, m	65.1, d ($^2J_{C,P}$ = 4 Hz)	4.10, m	64.9, d ($^2J_{C,P}$ = 4 Hz)

* ^{1}H (600 MHz) and ^{13}C (150 MHz) chemical shift values (δ) are given in ppm; ♯ pH = pD - 0.4; eq = equatorial, ax = axial.

5.5. REFERENCES

[1] S. A. Benner, A. M. Sismour, *Nat. Rev. Gen.* **2005**, *6*, 533-543.
[2] Y. Wu, M. Fa, E. L. Tae, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2002**, *124*, 14626-14630.
[3] D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Correa, F. E. Romesberg, *Nature* **2014**, *509*, 385-388.
[4] S. De, E. Groaz, P. Herdewijn, *Eur. J. Org. Chem.* **2014**, *2014*, 2322-2348.
[5] V. Pezo, F. W. Liu, M. Abramov, M. Froeyen, P. Herdewijn, P. Marlière, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 8139-8143.
[6] O. Prakash, Eisenber.Ma, *J. Bacteriol.* **1974**, *120*, 785-791.
[7] A. Piffeteau, M. Zamboni, M. Gaudry, *Biochim. Biophys. Acta* **1982**, *688*, 29-36.
[8] B. D. Davis, *Microbiol. Rev.* **1987**, *51*, 341-350.
[9] S. Magnet, J. S. Blanchard, *Chem. Rev.* **2005**, *105*, 477-498.
[10] H. W. Taber, J. P. Mueller, P. F. Miller, A. S. Arrow, *Microbiol. Rev.* **1987**, *51*, 439-457.
[11] B. Becker, M. A. Cooper, *ACS Chem. Biol.* **2013**, *8*, 105-115.
[12] A. Semenyuk, A. Földesi, T. Johansson, C. Estmer-Nilsson, P. Blomgren, M. Brännvall, L. A. Kirsebom, M. Kwiatkowski, *J. Am. Chem. Soc.* **2006**, *128*, 12356-12357.
[13] M. Sekine, S. Iimura, T. Nakanishi, *Tetrahedron Lett.* **1991**, *32*, 395-398.
[14] M. Manoharan, Y. Lu, M. D. Casper, G. Just, *Org. Lett.* **2000**, *2*, 243-246.
[15] N. W. Luedtke, P. Carmichael, Y. Tor, *J. Am. Chem. Soc.* **2003**, *125*, 12374-12375.
[16] I. Charles, L. Xue, D. P. Arya, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1259-1262.
[17] K. Michael, H. Wang, Y. Tor, *Bioorg. Med. Chem.* **1999**, *7*, 1361-1371.
[18] I. Hirao, K. Itoh, S. Nishino, Y. Araki, Y. Ishido, *Carbohydr. Res.* **1986**, *152*, 159-172.
[19] M. Suzuki, T. Sekido, S.-i. Matsuoka, K. Takagi, *Biomacromolecules* **2011**, *12*, 1449-1459.
[20] Q. Liu, N. W. Luedtke, Y. Tor, *Tetrahedron Lett.* **2001**, *42*, 1445-1447.
[21] V. A. Schepdael, R. Busson, P. Claes, *Bull. Soc. Chim. Bel.* **1992**, *101*, 709-718.
[22] M. Sainlos, P. Belmont, J.-P. Vigneron, P. Lehn, J.-M. Lehn, *Eur. J. Org. Chem.* **2003**, *2003*, 2764-2774.
[23] K. F. Blount, F. Zhao, T. Hermann, Y. Tor, *J. Am. Chem. Soc.* **2005**, *127*, 9818-9829.

Chapter 6: NUTRITIONAL EVALUATION METHODOLOGIES

6.1. Biological evaluation using auxotrophic bacterial strain

(Performed in collaboration with Lia Margamuljana and Elisabetta Groaz from Rega Institute and with scientist from Genoscope, Institut de Génomique, Evry, France)

In order to test the bacterial uptake of all synthetized peptide-dTMP conjugates **(2.19, 2.33-2.36, 2.38-2.42, 2.47-2.48, 2.57-2.58, 2.62, 2.65, 2.70** and **3.1, 3.2a, 3.2b, 3.15, 3.23, 3.38a, 3.38b, 3.4**), sulfono-dTMP phosphoramidate conjugates (**4.1-4.11**, **4.33** and **4.35**), biotin-dTMP conjugates (**5.7**, **5.10**, **5.13** and **5.15**) and kanamycin A-dTMP conjugates (**5.22** and **5.24**) as a source of TMP, an *E. coli* LN2666 *thyA*(-) strain[1] [a spontaneous streptomycin-resistant mutant of CB0129 (F^- W1485 *thiA*, *thyA*, *leu*, *deoB*, *rpsL*)[2] was employed. (*The strain was provided by Dr. Anirban Ghosh and Prof. Abram Aertsen, Centre for Food and Microbial Technology, Faculty of Bioscience Engineering, KU Leuven*). Its genotype is *E. coli* LN2666 Δ*phoA thyA(-)*. This strain is deleted for the *thyA* and *phoA* genes, which respectively encode for thymidylate synthase and alkaline phosphatase, in order to block all pathways that could lead to the formation or degradation of TMP inside the bacterial cells. All compounds were preventively analyzed (LC/MS or NMR) to exclude potential sample contamination, but no traces of thymine could be observed below detection limit. LB medium and LB agar were sterilized by autoclaving (120 °C, 30 min). Thymine (positive control), TMP (negative control) and all compounds were dissolved in Milli-Q water and sterilized through a 0.2μm sterile syringe filter (VWR). A single colony of the mutant strain was grown overnight at 37 °C in LB medium (3 mL). Cells were then harvested by centrifugation, washed and resuspended in fresh LB medium. The bacterial cells were then diluted 1:1000 in fresh LB medium before streaking on LB agar plates containing 20 μg/mL of each compound and grown overnight at 37 °C. Unfortunately, no colony formation was observed in the presence of compounds (peptide conjugates - **2.19, 2.33-2.36, 2.38-2.42, 2.47-2.48, 2.57-2.58, 2.62, 2.65, 2.70** and **3.1, 3.2a, 3.2b, 3.15, 3.23, 3.38a, 3.38b, 3.4)**, (sulfonate conjugates **- 4.1-4.11**, **4.33**, **4.35**), (biotin conjugates - **5.7**, **5.10**, **5.13**, **5.15** and kanamycin conjugates - **5.22**, **5.24)** or toxicity (kanamycin conjugates **5.22** and **5.24**) were observed.

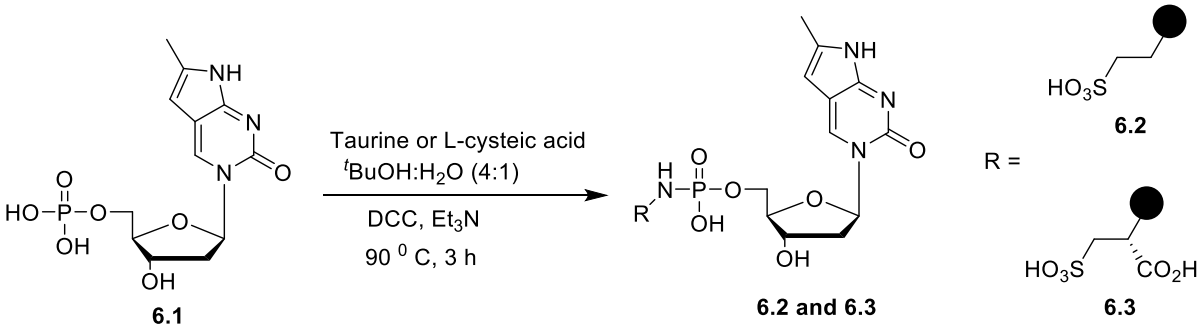
On the basis of preliminary biological screening none of the conjugates showed functional activity in *E.Coli* mutants lacking thymine. However, this might be due to several reasons, for instance high concentrations necessary for considerable growth, which are currently under further study. As additional evidences, biotin-TMP conjugates and sulfonate-TMP conjugates are undergoing biological evaluation using biotin and sulfonate auxotrophic *E. coli* mutants respectively.

6.2. Cell imaging with fluorescent microscopy (future work)

6.2.1. Synthesis of sulfono-phosphoramidate analogues containing a fluorescent base:

In parallel to the bacterial auxotrophic model, in order to gather additional evidences, we also planned to synthesize different biocarrier-nucleotide conjugates containing a fluorescent base, as uptake of these conjugates can be visualized and quantified using confocal microscopy through a live bacterial cell imaging. Although different fluorescent bases are available, pyrrolo-dC was selected for its non-toxic profile both for uptake and cell imaging purposes. Pyrrolo-dC mimics natural cytidine base (C) and its base pairing properties with guanidine (G) are not perturbed as it shows similar melting temperature (Tm) as natural C. Due to its small size it is also well accepted by most DNA and RNA polymerases and does not disturb the three dimensional structure of DNA. As the excitation and emission fluorescent spectra fall in the red zone (absorption band - 350 nm, emission band - 460 nm, ϕ_F = 0.2), visibility stands out from the noise of protein fluorescence and is significantly distinguishable from other natural nucleobases.

As a proof of concept, we first planned to synthesize two derivatives of this fluorescent nucleotide by linking taurine (2-aminoethanesulfonic acid) and L-cysteic acid through a phosphoramidate linker at the 5'-phosphate of the nucleotide. The synthesis of these phosphoramidate conjugates (Tau-PydCMP and L-Cys-PydCMP) was performed following a similar protocol described in **Chapter-4**, starting from pyrrolo-dC-5'-monophosphate triethylammonium salt and the corresponding sulfono-amine in presence of DCC. The uptake study of these conjugates (**6.2** and **6.3**) in *E. coli* and their evaluation with confocal fluorescent microscopy are being performed in collaboration with *Genoscope, Institut de Génomique, Evry, France.*



Pyrrolo-dC-5'-taurine phosphoramidate triethylammonium salt (6.2)

Pyrrolo-dC-5'-monophosphate triethylammonium salt (32 mg, 0.058 mmol) and 2-aminoethanesulfonic acid (21.9 mg, 0.175 mmol) were suspended in a 4:1 ^{t}BuOH/H_2O mixture (2 mL). Triethylamine (25 µL, 0.175 mmol) was then added to the suspension to facilitate dissolution, followed by DCC (48.2 mg, 0.234 mmol). The reaction mixture was heated at 90 ºC for 3 h. The reaction progress was monitored by TLC (iPrOH:H_2O:Et_3N 6:2:2). Upon completion, the reaction mixture was cooled to room temperature and the solvent was removed by rotary evaporation. The resulting crude material was purified by column chromatography on silica gel using the following gradient IPA:H_2O:NH_3 20:1:1, v/v/v; 15:1:1, v/v/v; 10:1:1, v/v/v, to provide

the desired nucleoside phosphoramidate as salt. Semi-preparative RP-HPLC (50 mmol TEAB in H_2O + 2% ACN and 50 mmol TEAB in 50% H_2O + 50% ACN) was employed for further purification. The residue was lyophilized and freeze-dried repeatedly until constant mass to obtain the triethylammonium salt of compound **6.2** (27.2 mg, 71%) as a white solid. ^{1}H NMR (300 MHz, D_2O) δ = 8.52 (s, 1H, H-9), 6.33 (app t, J = 6.3 Hz, 1H, H-1'), 6.07 (d, J = 1.2 Hz, 1H, H-7), 4.50-4.45 (m, 1H, H-3'), 4.19-4.18 (m, 1H, H-4'), 4.07-3.92 (m, 2H, H-5' and H-5''), 3.15-3.05 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 2.97-2.90 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H, mergerd with Et$_3$N), 2.57-2.22 (m, 2H, H-2' and H-2''), 2.21 (s, 3H, CH_3-Pyrrolo-dC); ^{13}C NMR (75 MHz, D_2O) δ = 157.4 (C-4), 155.3 (C-2), 139.7 (C-6), 134.6 (C-9), 111.3 (C-8), 97.8 (C-7), 87.5 (C-1'), 85.8 (d, $^3J_{C,P}$ = 8.9 Hz, C-4'), 70.2 (C-3'), 63.3 (d, $^2J_{C,P}$ = 5.0 Hz, C-5'), 52.0 (d, $^3J_{C,P}$ = 6.5 Hz, CH$_2$C$_\beta$H$_2$SO$_3$H), 40.2 (C-2'), 36.7 (C$_\alpha$H$_2$CH$_2$SO$_3$H), 12.4 (CH$_3$-Pyrrolo-dC); ^{31}P NMR (121 MHz, D_2O) δ = 8.1; HRMS for $C_{14}H_{21}N_4O_9PS$ $[M-H]^-$ calcd.: 451.0696, found: 451.0694.

Pyrrolo-dC-5'-(L-cysteic acid) phosphoramidate triethylammonium salt (6.3)

The triethylammonium salt of compound **6.3** was obtained as an off-white solid (77.1 mg, 24%) following the same procedure as the one used for compound **6.2**, starting from pyrrolo-dC-5'-monophosphate triethylammonium salt (220 mg, 0.402 mmol), L-cysteic acid (204 mg, 1.205 mmol), triethylamine (0.168 mL, 1.205 mmol), DCC (331.5 mg, 1.607 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (10 mL) at 90 ºC for 3 h. ^{1}H NMR (500 MHz, D_2O) δ = 8.63 (s, 1H, H-9), 6.40 (app t, J = 6.3 Hz, 1H, H-1'), 6.17 (s, 1H, H-7), 4.56-4.54 (m, 1H, H-3'), 4.23-4.22 (m, 1H, H-4'), 4.12-4.04 (m, 2H, H-5' and H-5''), 3.91-3.86 (m, 1H, CHCO$_2$H), 3.23-3.19 (m, 2H, CH_2SO$_3$H), 2.58-2.30 (m, 2H, H-2' and H-2''), 2.28 (s, 3H, CH_3-Pyrrolo-dC); ^{13}C NMR (125 MHz, D_2O) δ = 178.3 (d, $^3J_{C,P}$ = 3.6 Hz, CO$_2$H), 157.0 (C-4), 155.0 (C-2), 139.2 (C-6), 134.6 (C-9), 111.2 (C-8), 97.8 (C-7), 87.0 (C-1'), 85.6 (d, $^3J_{C,P}$ = 8.6 Hz, C-4'), 69.7 (C-3'), 62.9 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 54.1 (d, $^3J_{C,P}$ = 6.9 Hz, CH$_2$SO$_3$H), 53.4 (CHCO$_2$H), 39.7 (C-2'), 12.1 (CH$_3$-Pyrrolo-dC); ^{31}P NMR (202 MHz, D_2O) δ = 6.1; HRMS for $C_{15}H_{21}N_4O_{11}PS$ $[M-H]^-$ calcd.: 495.0592, found: 495.0596.

6.3. References:

[1] F. Cornet, J. Louarn, J. Patte, J. M. Louarn, *Genes & Development* **1996**, *10*, 1152-1161.

[2] R. E. Bird, J. Louarn, J. Martuscelli, L. Caro, *Journal of Molecular Biology* **1972**, *70*, 549-566.

Chapter 7: General Discussion and Perspectives

In the last two decades research efforts have been devoted to the development of xeno nucleic acids (XNAs) through modifications involving sugar backbone, base moiety or phosphodiester linkage. The role of non-natural nucleic acids has been rigorously studied for their potential application as therapeutics, in material sciences, molecular diagnostics, biotechnology and life sciences (antigene, antisense, siRNAs, cellular probes, aptamers, biosensors and XNAzymes). XNAs bearing base modifications, mimicking the canonical and non-canonical base pairing systems, have also been examined in the context of the genetic alphabet. Over the last few years, research aiming to achieve backbone modifications by replacing the (deoxy)ribose sugar, has clearly established that synthetic variants of natural polymers are also capable of storing genetic information.

With the advent of solid phase synthesis of sequence-defined gene fragments and DNA-oligomers, along with modern chemical biology methods like sequence assembling, PCR and recombinant technology, synthetic genes of interest have started to be evaluated in the reprogramming of living entities for a variety of applications in energetics, agriculture and bio-technology. However, risks of genetic pollution of the natural ecosystem are possible due to deliberate or unintended damage arising from genetic information exchanging (vertical or horizontal gene transfer) between synthetic genetic constructs and natural habitats, a major concern of biosafety in the society. To address such issues, the introduction of new technologies to avoid or limit genetic cross talk has been pursued, together with the design of artificial genetic constructs that could be safely used for beneficial purposes.

Recently, it was hypothesized that the use of an artificial orthogonal system, a third type of self-reliant information system other than DNA and RNA, could serve the purpose of avoiding genetic cross-talk and pollution with the natural complements.[1] Modified nucleic acids or Xeno-nucleic acid (XNAs) were therefore designed and synthesized. In spite of the advancement in the chemical synthesis of nucleic acid oligomers, XNAs synthesis is often limited to <100 nucleotide. Additionally, enzymatic polymerization of XNAs with high fidelity remains a huge challenge due to the low selectivity of XNA building blocks in the active pocket of natural polymerase used for this purpose. The directed evolution of the active site of different enzymes using advanced techniques like phage display, CSR or CST is highly desired, for catalysing different XNA based biological functions with high selectivity, efficiency and fidelity. So far, Holliger and co-workers designed and fabricated engineered polymerases which can efficiently recognize, incorporate and transcribe sugar modified

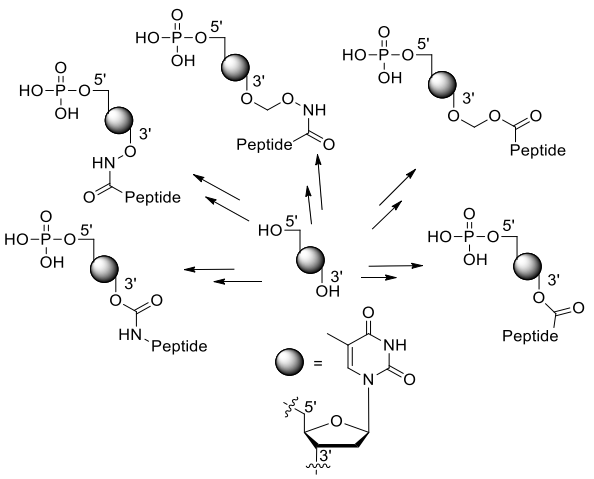
XNAs (HNA, CeNA, ANA, FANA, TNA, and LNA) on a DNA based template and are also able to reverse-transcribe to DNA on XNA based templates. This represents a remarkable contribution to the field of XNA based directed enzyme evolution for *in vitro* assembling, replication and propagation of XNAs.[2] Very recently, the *in vitro* functional activity of XNAzymes that were devised on four XNA backbones (ANA, FANA, HNA and CeNA), mimicking RNA and DNA based enzymes was also successfully demonstrated.[3] Both of their achievements are a milestone for the establishment of alternative genetic polymers (XNAs) for their abilities to endure heredity and evolution.

Because of the relatively simple genetic architecture of prokaryotes, the first in vivo invasion with XNA replicons was thought to be more likely to happen in bacteria. In order for the reprogrammed bacterial cells to reach their true evolutionary potential, the controlled *in vivo* propagation of XNA polymers is essential, and firstly requires the availability of XNA precursors in sufficient amount inside the bacterial cells. As XNA precursors do not occur in nature, their corresponding active intermediates of (XNA nucleotides) need to be supplied from outside the cells. Although most modified nucleosides can easily enter the bacterial cell, they are poorly recognized and phosphorylated to their monophosphate form by the host nucleoside kinases. On the other hand, modified nucleotides are charged and highly polar at physiological pH and their transport inside bacterial cells is restrained by the hydrophobic lipid bi-layer present in the bacterial cell wall. To solve this issue, so far efforts have involved over-expression of endogenous nucleoside kinases e.g. from *Drosophila Melanogaster* with broad substrate specificity in *E. coli*,[4] which can provide modified nucleoside monophosphates ensuing a nucleoside salvage pathway, and lately by algal nucleoside triphosphate transporters in bacterial cells which efficiently facilitates active transport of triphosphates of d5SICS and dNaM.[5] Selectivity and recognition of XNA-triphosphates by different algal triphosphate transporters represented a significant contribution towards the establishment of the first XNA-nucleotide uptake system and *in vivo* replication of XNA base pairs in bacterial cells, but broad substrate specificity for all type of XNA-triphosphates is still a huge concern of this approach.

On the other hand, bacteria express different transmembrane transport proteins for the uptake of different nutrients against a concentration gradient, for their cellular survival such as amino acids, peptides, ions, sugars, vitamins like biotin and thiamine, sulfates and sulfonates, and other hydrophilic molecules, which otherwise would have a restricted entry through the rigid lipid bi-layer of the cell wall of bacteria. By exploiting such active transport systems it may be possible to deliver different cargo inside the bacterial cells, as it was

demonstrated utilizing peptide transport systems for delivery of different antimicrobial[6] and therapeutic[7] compounds with the enhanced efficacy compared to direct means. Likewise, the delivery of large peptides (up to 31-mer) was also promoted by biotinylation, which otherwise was not possible or limited to five to six-mer for non-biotinylated ones.[8] Interestingly, these types of natural delivery systems have never been explored for nucleotide delivery inside the bacterial cells.

Therefore, this thesis is mainly aimed at the systematic exploration of different types of natural transporters and uptake systems in an attempt to facilitate the delivery of nucleotides inside bacterial cells, based on the synthesis and evaluation of a wide range of nucleotide-conjugates through a nutritional selection approach. For this purpose, thymidine mono-phosphate (TMP) was chosen as an exemplary nucleotide in a model study. Based on this hypothesis, an auxotrophic *E.coli* strain deleted for the ThyA gene encoding for thymidylate synthase was constructed. TMP-conjugates are then provided in extracellular medium to such strain which absolutely require thymine or thymidine for growth. The viability of this model can be visualized from the growth of thymidine auxotrophic *E.coli* strain in presence of TMP-conjugates, in a condition only when TMP-conjugate is taken up and TMP is released as nutrient inside the cells. Diverse conjugation strategies involving a covalent linker between either the 3'-hydroxyl or the 5'-phosphate position of TMP and selected delivery vehicles were designed such a way that linkers should be chemically or enzymatically cleavable and release the attached cargo, once internalized inside the cells. As additional evidence of this model, the synthesis of different conjugates containing a non-toxic fluorescent base mimicking the natural base and their evaluation under confocal fluorescent microscopy were also planned.



Linker	Peptide
OCONH	For-M-K-OMe, For-G-K-F-OMe, For-G-βA-F-OMe, For-M-K-F-OMe, For-M-K-A-OMe, For-M-K-K-OMe, For-M-K-A-F-OMe, For-M-K-A-A-F-OMe, A-K-A-OMe, A-K-A
ONHCO	For-M-E-OMe
OCH_2ONHCO	For-M-E-OMe, For-M-E-NH_2
OCH_2OCO	For-G-E-A-OMe
OCO	For-M-K-K, For-G-E-F

Figure 7-1. Chemical structure of novel peptide-nucleotide conjugates (PNCs).

In **Chapter 2**, a series of peptide-TMP conjugates were designed and synthesized to evaluate their potential as delivery systems of nucleotides uptake in an auxotrophic *E. coli* strain. A variety of di- to penta-peptides was chosen as delivery vehicles mainly targeting different peptide permeases (Dpp, Tpp, Opp). As free N- and C-terminus and basic peptide residues are important for recognition by periplasmic binding proteins, whilst the protection of the N- and C-terminus could avoid extracellular degradation, peptidic sequences combining those features were accordingly designed. Different bioreversible linkages such as carbamate (OCONH), ester (OCO), oxyamide (ONHCO), oxymethyleneoxyamide (OCH_2ONHCO), oxymethylene-oxyester (OCH_2OCO) were introduced to allow conjugation of side chain functionalities of oligopeptides to the 3'-hydroxyl group of TMP, to tune metabolic stability. In this chapter we discussed different linear and convergent synthetic strategies in detail for the preparation of peptide-nucleotide conjugates (PNCs) along with orthogonal protection-deprotection chemistries starting from thymidine, as summarized in Fig. 7-1. Notably, the protection of carboxylic acid groups was not necessary with an incoming amino acid in the coupling step, during oligopeptide synthesized using NHS-DCC active ester method in solution phase. The construction of PNCs were accomplished either by conjugation of the peptide chain at the 3'-position of thymidine followed by phosphorylation at 5'-position or vice versa with respect to reactions order. Although attempts to synthesize PNCs with carbamoyl (OCONH) linker via the first approach was successful but it was not suitable for the synthesis of methionine containing peptides, where selective oxidation of P^{III} to P^{V} was challenging in presence of sulfur. Such compounds could however be obtained following the second approach. The key steps for the preparation of PNCs with an oxyamide linker were the formation of 5'-protected *xylo*-thymidine followed by Mitsunobu inversion with a protected hydroxyl amino functionality. Very mild condition using diluted methyl amine (4%) was chosen for phthalimide deprotection to avoid side reactions of PNCs bearing delicate functionalities. Pummerer rearrangement and further functionalization of the thiomethyl group were the crucial steps for the synthesis of PNCs with oxymethyleneoxyamide, oxymethyleneoxyester linkers. Final deprotection of all conjugates was realized in high yield via hydrogenolysis, using either 10% Pd/C or 20% $Pd(OH)_2$ catalysts, to avoid over reduction of the thymine base. To our surprise, the thiomethyl functionality present in methionine-containing PNCs wad stable to such conditions. On the basis of a preliminary biological screening none of the conjugates showed colony growth in the presence of a thymidine auxotrophic *E.coli* strain, which may arise from different factors like requirement of higher concentrations or metabolic (in)stability of conjugates. With a

view to the development of a xenonucleotides delivery system in bacterial cells, the introduction of other types of residues may be explored in future, based on these peptide conjugation chemistries.

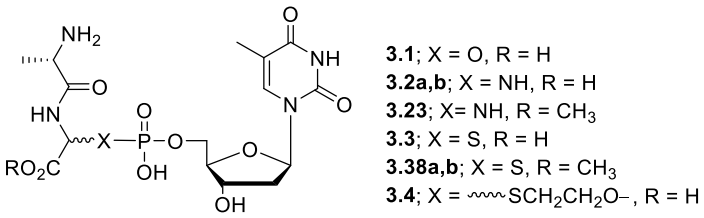


Fig 7.2: Structures of α-substituted glycine peptide-nucleotide conjugates with different phospho-ester linkages.

Aiming at targeting peptide delivery systems, α-substituted glycine containing peptides were successfully exploited to transport and liberate cargos such as antimicrobial agents or toxophoric agents inside bacterial cells following the metabolic action of intracellular peptidases. In **Chapter 3** we described the design and synthesis of various α-substituted glycine PNCs with TMP. All the PNCs derived from glycine dipeptides were linked through the α-carbon of the C-terminal glycine residue to the 5'-hydroxyl of thymidine through various phosphoester linkers i.e. phosphate, phosphoramidate, phosphorothioate and thioethyl-phosphate. Synthesis of such conjugates were accomplished either by nucleophilic substitution at the α-position of glycine or at the phosphorous-group itself, as anticipated from retro-synthetic analysis. Following the second route for the synthesis of conjugate **3.1** with a phosphate linkage, attempts were made using different phosphorous activation pathways like phosphoramidite or $POCl_3$ based approaches, but met with failure presumably because of low nucleophilicity at the α-hydroxyl functionality of glycine. By inversing the reaction order and using the phosphate as nucleophile and OAc as leaving group, synthesis of **3.1** was successful although low yields were obtained due to side products formation arising from competitive nucleophilic attack of the NH-group of the thymine base. The phosphoramidate analogues **3.2a,b** were prepared following DCC-mediated coupling between TMP and an α-amino glycine derivative. The separation of diastereomers **3.2a** and **3.2b** easily accomplished with preparative-HPLC in the pre-final step with protected amino acids. The synthesis of thioethyl-phosphate derivative **3.4** was realized in good yield by altering the protecting groups at the C-and N-terminal of peptide chain and at 3'-position of thymidine with orthogonal protecting groups like OMe, Boc and OBz respectively, due to the

problematic hydrogenolysis of the Cbz and OBn groups in the final step, arising from catalyst poising by the sulphur-atom present. The isolation and purification of thioester derivative **3.3** were challenging, but methyl ester derivative **3.38a** and **3.38b** were synthesized and purified following a 1,3,2-oxathiaphospholane approach. Generally, the α-substituent in α-substituted glycines acts as a good leaving group, and in addition to that anchimeric assistance of free carboxylic groups from peptide makes the phosphorous-linker more vulnerable to hydrolytic cleavage. It was assumed that protection of the free carboxylic acid with an ester group would increase the stability of conjugates, as demonstrated by the synthesis and relative stability studies of methyl ester derivatives **3.23**. The biological profile of all the conjugates was evaluated by *in vitro* DNA chain elongation experiments, catalysed by different thermophilic and mesophylic microbial DNA polymerases, comprising Therminator, Vent (exo-) and the Klenow fragment (exo-) of Pol I. Only compounds **3.2a** and **3.2b** showed good incorporation properties with the Klenow fragment indicating that a free carboxylic group is necessary for binding at the active site of the enzyme. The lower molecular mechanics energy of **3.2a** over **3.2b** with this enzyme explains the better fidelity of **3.2a**, as indicated from a homology model in a molecular modeling study. The absolute configurations of **3.2a** and **3.2b** were also determined based on the same model. Kinetic parameter showed that the V_{max}/K_m of **3.2a** is only 994-fold lower than the natural substrate (TTP), comparable to previously explored alternative leaving groups in our lab. Although in the uptake assay with a thymidine auxotrophic *E.coli* strain none of these conjugates showed functional activity, in view of the good incorporation properties of phosphoramidates analogues, future efforts will focus on broadening the scope of substrates linked at the 5'-position of nucleotides with different linkers to establish an optimal stability for binding to periplasmic proteins along with improved physico-chemical properties in the DNA polymerisation reaction.

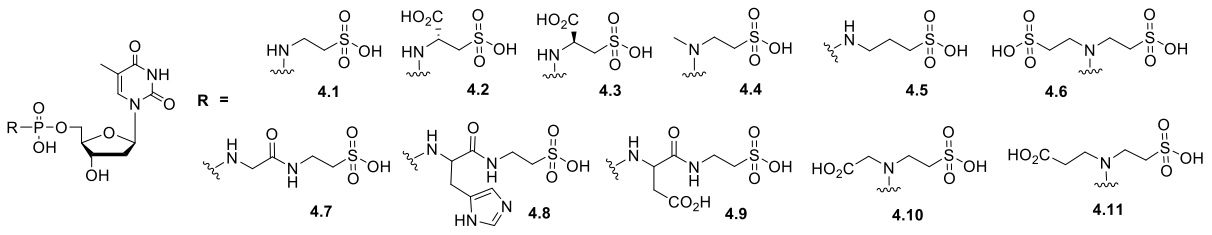


Figure 7-3. Overview of sulfonate-containing phosphoramidate analogues of thymidine.

In **chapter 4**, several novel modified phosphono-5'-phosphoramidate nucleosides have been synthesized featuring metabolites such as taurine and a broad range of aliphatic sulfonates (summarized in Fig.7-3) coupled through a P-N bond to the 5′-phosphate position of

deoxynucleotides. The design of different sulfono-groups in conjugates aimed both at exploitation of sulfonate transport systems for nucleotide uptake and at the development of alternative leaving groups capable of being processed by DNA polymerases for multiple nucleotide incorporations mimicking the pyrophosphate group. Synthesis of all the conjugates was achieved by different methods including phosphoramidate coupling like DCC mediated coupling and the imidazolide approach. Remarkably, in an improved protocol the protection of the sulfonic acid group was not necessary. Both thermophilic and mesophylic microbial polymerases Vent(exo-), Therminator and the Klenow fragment were used for *in vitro* template-directed DNA polymerization studies. Amongst them, taurine-N-acetic acid-dTMP showed good substrate property, therefore such phosphoramidate derivatives with all four nucleobases were synthesized and evaluated to establish the influence of the base moiety on substrate properties. Results showed that the incorporation efficacy followed the order G > A ≥ T > C. Therminator polymerase exhibited the broader substrate-specificity but best incorporation results were obtained with most effective polymerase Klenow fragment (exo-) and taurine-N-acetic acid-dGMP (**4.28)** as substrate. A kinetic study revealed that the V_{max}/K_m ratio of compound **4.28** was 702-fold lower than the natural substrate dGTP, within the range of previously explored bio-isosteric leaving groups (IDA, IDP and L-Asp). Unfortunately, none of this thymidine containing conjugates was able to sustain the growth of thymidine auxotrophic *E. coli* mutants. Therefore, it is important to investigate and broaden the substrate scope for this series of compounds to find structural similarities that can be relevant at the same time for the developments of substrates for DNA polymerases functional at room temperature, and for recognition by sulfonate transport systems through the modulation of physical parameters (e.g. polarisibility, solubility, binding constant, pKa, size and spatial arrangement). Apart from their interest in nucleotide delivery, phosphate group modified nucleotides discussed in **Chapter 3** and **4** could be of use for other applications such as the polymerase chain reaction (PCR) for *in vitro* construction of DNA or RNA fragments or as pro-drugs for antiviral therapy.

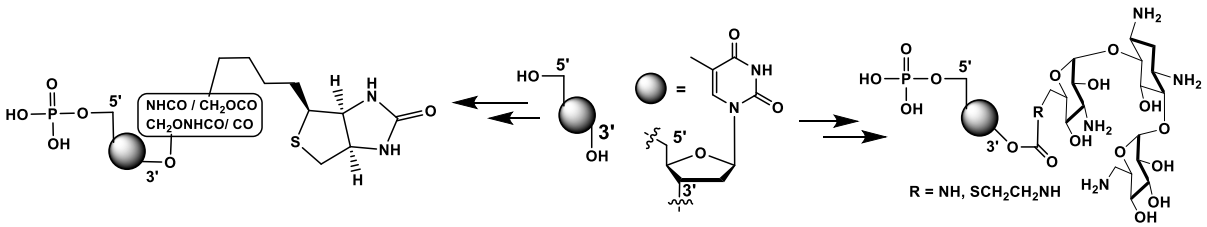


Fig 7-4: Structures of nucleotide conjugates of biotin and kanamycin with different linkages.

In **Chapter 5**, a series of TMP conjugates of biotin and kanamycin A (Fig. 7-4) were designed and synthesized in the search for a nucleotide delivery system. Biotin-TMP

conjugates were linked through the carboxyl group of biotin to the 3'-OH group of TMP featuring an oxymethyleneoxy ester, a carboxy ester, a carbamate and oxymethyleneoxyamide linker, adopting previously optimized conjugation strategies such as coupling and debenzylation procedures used for PNCs (**chapter 2**). Similarly, kanamycin A-TMP conjugates were constructed in a convergent manner with a carbamoyl or thioethyl carbamoyl group, exploiting the different chemical reactivity between the primary hydroxyl and amino groups amongst polyhydroxyl and polyamino functionalities present in the structure of kanamycin A. 2D and variable-temperature NMR spectroscopic analysis established the alignment of the carbamoyl linkage and the formation of rotamers, which may be arising from an inherent flexibility around glycosidic bonds for the kanamycin-TMP conjugates. On the basis of biological evaluation using the auxotrophic model, none of the biotin-TMP or the kanamycin A-TMP conjugates was able to sustain growth of a thyA-, ΔphoA *E. coli* strain. However, based on these synthetic strategies of conjugation several other vitamin-TMP and polycationic-TMP conjugates could be synthesized and screened.

In summary, through the systematic modification of synthetic TMP conjugates obtained by introduction of biocleavable linkers either at the 3'- or 5'-position of TMP through different metabolite hiring from microbial transport systems, in this thesis we have expanded the chemical scope of the potential vectors for the active uptake of nucleotides in bacterial cells. We have successfully demonstrated novel synthetic strategies to achieve chemical conjugation via linear and convergent methods together with orthogonal protection-deprotection sequences. The 5'-modified nucleotide conjugates were designed in view of their intracellular metabolism as substrates for DNA polymerase and some of them showed moderate to good substrate properties. Unfortunately, it is apparent form preliminary biological data, that none of the vectors was successful for actively delivering TMP inside the bacterial cell. Although none of the conjugates was capable to produce growth of a thymidine auxotrophic LN2666 *E.coli* strain (thyA-, ΔphoA) (**Chapter 6**), this can be rationalized as due to numerous parameters e.g. (1) product concentration needed and type of culture media (2) mode of addition e.g. continuous or single shot (3) stability of conjugates in extracellular media (4) phosphatase activity other than phoA (5) binding constant with periplasmic binding protein (6) recognition and transportation through permeases (7) desired intracellular metabolism of conjugates releasing TMP or 5'-modified conjugates as substrate for *E. coli* DNA polymerase Pol I.

To gain supplementary proof in a nutritional model, peptide (methionine)-TMP, biotin-TMP and sulfonate-TMP conjugates are currently undergoing biological evaluation using methionine, biotin and sulfonate auxotrophic *E. coli* mutants respectively. Fluorescent conjugates Tau-PydCMP and L-Cys-PydCMP are also being evaluated through live cell imaging techniques under confocal fluorescent microscopy. Recent preliminary data showed of sulfonate-TMP and biotin-TMP conjugates to be positive in sulphur-auxotrophic and biotin-auxotrophic strains respectively, but further investigations are needed to confirm whether this is caused by extracellular degradation, as bacteria require sulphur and biotin in lower concentrations than TMP for survival.

As nucleotides delivery in bacterial cells remains a challenging issue, we envisioned that future research should mainly emphasise on the following areas:

1. To increase the chemical space between the conjugation partners with the variation and introduction of other types of unexplored bio-labile linkers and the use of delivery units based on other nutritional selection approaches e.g. sugars and vitamins etc.
2. To perform stability studies of different conjugates with advanced techniques that can detect metabolites in low concentration e.g. HPLC or LCMS, in order to tune metabolic stability and unambiguously clarify biological data resulting from different auxotrophic models.
3. To modify the 5'-phosphate group with neutral or hydrophobic moieties to increase the feasibility of passive diffusion with the aim of intracellular release of nucleoside mono phosphate or a metabolite that can be a substrate for DNA polymerases. Masking of the two anionic charges arising from naked 5'-phosphate groups is thought to influence the hydrophobic interactions with the lipid bi-layer and the susceptibility to the extracellular phosphatases degradation.
4. To gain an understanding to the mode of transport of the various conjugates, the recognition and binding studies of different periplasmic binding proteins and permeases are very important, aided by *in silico* molecular modeling and different bio-physical assays. Besides, over-expression of a particular transporter during uptake study based on a type of conjugate will also shed a light on the amount of TMP needed for auxotrophic bacterial growth.
5. To develop and establish double auxotrophic (e.g. thymidine and methionine, thymidine and biotin, thymidine and taurine) bacterial mutants along with the deletion of all genes encoding for enzymes with phosphatase activity, would be an ideal

auxotrophic model to establish nucleotide uptake even at low substrate concentration, in parallel to the use of non-toxic fluorescent bases.

In short, it is once again significant to remember that delivery of nucleotides inside bacterial cells is essential for the construction and evolution of orthogonal genetic systems along with systematic exploration of nucleobases, sugar backbones, leaving groups and engineered polymerases to address important bio-safety issues. Moreover, the successful identification of a natural transport system able to deliver nucleoside mono-phosphate (NMP) inside the bacterial cells could be further exploited for delivery of different XNA-nucleotides and would aid the search for nucleotide-mimetic antibacterial drugs that are inefficient *in vivo* due to poor cell penetrating ability. As no nucleoside monophosphate transporter is known in the wild ecosystem to date, the only viable way up to now to actively transport nucleotides is by over-expressing algal nucleotide triphosphate (NTP) transporters that showed actively transport XNA-triphosphates in *E. coli*. However this approach was found to be substrate dependent with varied specificity and rate of transportation, as demonstrated by Romesberg and co-workers recently.[5] Moreover over-expression of NTP-transporters and production of NTPs are not economic from a synthetic point of view. Our approach to nucleotide delivery is thought to be much simpler, as it relies on the exploitation of a delivery system already existing in bacteria.

References:

[1] P. Herdewijn, P. Marliere, *Chemistry & Biodiversity* **2009**, *6*, 791-808.
[2] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S. Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, P. Holliger, *Science* **2012**, *336*, 341-344.
[3] A. I. Taylor, V. B. Pinheiro, M. J. Smola, A. S. Morgunov, S. Peak-Chew, C. Cozens, K. M. Weeks, P. Herdewijn, P. Holliger, *Nature* **2015**, *518*, 427-430.
[4] Y. Wu, M. Fa, E. L. Tae, P. G. Schultz, F. E. Romesberg, *Journal of the American Chemical Society* **2002**, *124*, 14626-14630; S. K. Jarchow-Choy, E. Sjuvarsson, H. O. Sintim, S. Eriksson, E. T. Kool, *Journal of the American Chemical Society* **2009**, *131*, 5488-5494.
[5] D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Correa, F. E. Romesberg, *Nature* **2014**, *509*, 385-388.
[6] S. M. Hammond, A. Claesson, A. M. Jansson, L. G. Larsson, B. G. Pring, C. M. Town, B. Ekstrom, *Nature* **1987**, *327*, 730-732.
[7] T. E. Fickel, C. Gilvarg, *Nat New Biol* **1973**, *241*, 161-163.
[8] J. R. Walker, E. Altman, *Applied and Environmental Microbiology* **2005**, *71*, 1850-1855.

CV and publications

Swarup De was born in 1983 in the Bankura district of West Bengal, India. He studied his Bachelor of Science (B. Sc.) in Chemistry (honours) in Bankura Sammilani College under Burdwan University, India from 2000-2003. Then he moved to department of Chemistry, Indian Institute of Technology Kharagpur (IIT-KGP), India from 2003-2005, where he obtained his Master of Science (M.Sc.) in Chemistry under supervision of Prof. Dr. J. K. Ray. He also worked as a Synthetic Organic Chemist for Chembiotek and Syngenta in India for five years before to start his doctoral study in Sept 2011 at KU Leuven under the supervision of Prof. Dr. Piet Herdewijn in the Laboratory of Medicinal Chemistry, Rega Institute for Medical Research, Belgium.

Publications:

1. ***Swarup De****, Elisabetta Groaz and Piet Herdewijn.** Tailoring Peptide–Nucleotide Conjugates (PNCs) for Nucleotide Delivery in Bacterial Cells. *Eur. J. Org. Chem*., **2014**, *11*, 2322–2348

2. ***Swarup De,*** *Elisabetta Groaz, Mohitosh Maiti, Valerie Pezo, Philippe Marliere and Piet Herdewijn.** Synthesis of new biocarrier-nucleotide systems for cellular delivery in bacterial auxotrophic strains. *Tetrahedron* **2014,** *70*, 8843-8851.

3. ***Swarup De***, *Elisabetta Groaz, Mikhail Abramov, Lia* Margamuljana*, Philippe Marliere and Piet Herdewijn.** Sulfonate derived phosphoramidates as active intermediates in the enzymatic primer-extension of DNA. *Org. Biomol. Chem.* **2015**, *13*, 3950-3962

4. ***Swarup De***, *Elisabetta Groaz, Lia Margamuljana, Mathy Froeyen and Piet Herdewijn.** 5'-linked peptide nucleotide conjugates for nucleoitdes delivery in bacterial cells. **2015** *Manuscript under preparation.*